APPLICATIONS OF STEIN'S METHOD FOR CONCENTRATION INEQUALITIES

SOURAV CHATTERJEE AND PARTHA S. DEY

Abstract. Stein's method for concentration inequalities was introduced to prove concentration of measure in problems involving complex dependencies such as random permutations and Gibbs measures. In this paper, we provide some extensions of the theory and three applications: (1) We obtain a concentration inequality for the magnetization in the Curie-Weiss model at critical temperature (where it obeys a non-standard normalization and super-Gaussian concentration). (2) We derive exact large deviation asymptotics for the number of triangles in the Erdős-Rényi random graph $G(n,p)$ when $p \geq 0.31$. Similar results are derived also for general subgraph counts. (3) We obtain some interesting concentration inequalities for the Ising model on lattices that hold at all temperatures.

1. Introduction

In his seminal 1972 paper [36], Charles Stein introduced a method for proving central limit theorems with convergence rates for sums of dependent random variables. This has now come to be known as *Stein's method*. The technique is primarily used for proving distributional limit theorems (both Gaussian and non-Gaussian). Stein's attempts [37] at devising a version of the method for large deviations did not prove fruitful. Some progress for sums of dependent random variables was made by Raič [35]. The problem was finally solved in full generality in [10]. A selection of results and examples from [10] appeared in the later papers [11, 12]. In this paper we extend the theory and work out three further examples. The paper is fully self-contained.

The sections are organized as follows. In Section 2 we state the main results, the examples, and some proof sketches. The complete proofs are in Section 3.

2. Results and examples

The following abstract theorem is quoted from [11]. It summarizes a collection of results from [10]. This is a generalization of Stein's method of exchangeable pairs to the realm of concentration inequalities and large deviations.

Theorem 1 ([11], Theorem 1.5)**.** *Let $\mathcal{X}$ be a separable metric space and suppose (X, X') is an exchangeable pair of $\mathcal{X}$-valued random variables. Suppose $f : \mathcal{X} \to \mathbb{R}$ and $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are square-integrable functions such that F is antisymmetric (i.e. $F(X, X') = -F(X', X)$ a.s.), and $\mathbb{E}(F(X, X') \mid X) = f(X)$ a.s. Let*

$$\Delta(X) := \frac{1}{2}\,\mathbb{E}\big(|(f(X) - f(X'))F(X, X')| \,\big|\, X\big).$$

Then $\mathbb{E}(f(X)) = 0$, and the following concentration results hold for $f(X)$:

2000 *Mathematics Subject Classification.* Primary: 60E15, 60F10; Secondary: 60C05, 82B44.

Key words and phrases. Stein's method, Gibbs measures, Concentration inequality, Ising model, Curie-Weiss model, Large deviation, Erdős-Rényi random graph, Exponential random graph.

The research was supported in part by NSF grant DMS-0707054 and a Sloan Research Fellowship.

(*i*) *If* $\mathbb{E}(\Delta(X)) < \infty$, *then* $\mathrm{Var}(f(X)) = \frac{1}{2}\,\mathbb{E}((f(X) - f(X'))F(X, X'))$.

(*ii*) *Assume that* $\mathbb{E}(e^{\theta f(X)}|F(X, X')|) < \infty$ *for all* θ. *If there exists nonnegative constants* B *and* C *such that* $\Delta(X) \leq Bf(X) + C$ *almost surely, then for any* $t \geq 0$,

$$\mathbb{P}\{f(X) \geq t\} \leq \exp\left(-\frac{t^2}{2C + 2Bt}\right) \quad \text{and} \quad \mathbb{P}\{f(X) \leq -t\} \leq \exp\left(-\frac{t^2}{2C}\right).$$

(*iii*) *For any positive integer* k, *we have the following exchangeable pairs version of the Burkholder-Davis-Gundy inequality:*

$$\mathbb{E}(f(X)^{2k}) \leq (2k - 1)^k\, \mathbb{E}(\Delta(X)^k).$$

Note that the finiteness of the exponential moment for all θ ensures that the tail bounds hold for all t. If it is finite only in a neighborhood of zero, the tail bounds will hold for t less than a threshold.

One of the contributions of the present paper is the following generalization of the above result for non-Gaussian tail behavior. We apply it to obtain a concentration inequality with the correct tail behavior in the Curie-Weiss model at criticality.

Theorem 2. *Suppose* (X, X') *is an exchangeable pair of random variables. Let* $F(X, X'), f(X)$ *and* $\Delta(X)$ *be as in Theorem* 1. *Suppose that we have*

$$\Delta(X) \leq \psi(f(X)) \text{ almost surely}$$

for some nonnegative symmetric function ψ *on* $\mathbb{R}$. *Assume that* ψ *is nondecreasing and twice continuously differentiable in* $(0, \infty)$ *with*

$$\alpha := \sup_{x>0} x\psi'(x)/\psi(x) < 2 \tag{1}$$

$$\text{and } \delta := \sup_{x>0} x\psi''(x)/\psi(x) < \infty. \tag{2}$$

Assume that $\mathbb{E}(|f(X)|^k) < \infty$ *for all positive integer* $k \geq 1$. *Then for any* $t \geq 0$ *we have*

$$\mathbb{P}(|f(X)| > t) \leq c \exp\left(-\frac{t^2}{2\psi(t)}\right)$$

for some constant c *depending only on* α, δ. *Moreover, if* ψ *is only once differentiable with* $\alpha < 2$ *as in* (1), *then the tail inequality holds with exponent* $t^2/4\psi(t)$.

An immediate corrolary of Theorem 2 is the following.

Corollary 3. *Suppose* (X, X') *is an exchangeable pair of random variables. Let* $F(X, X'), f(X)$ *and* $\Delta(X)$ *be as in Theorem* 1. *Suppose that for some real number* $\alpha \in (0, 2)$ *we have*

$$\Delta(X) \leq B\,|f(X)|^\alpha + C \text{ almost surely}$$

where $B > 0, C \geq 0$ *are constants. Assume that* $\mathbb{E}(|f(X)|^k) < \infty$ *for all positive integer* $k \geq 1$. *Then for any* $t \geq 0$ *we have*

$$\mathbb{P}(|f(X)| > t) \leq c_\alpha \exp\left(-\frac{1}{2} \cdot \frac{t^2}{Bt^\alpha + C}\right)$$

for some constant c_α *depending only on* α.

The result in Theorem 2 states that the tail behavior of $f(X)$ is essentially given by the behavior of $f(X)^2/\Delta(X)$. Condition (1) implies that $\psi(x) < \psi(1)(1 + x^2)$ for all $x \in \mathbb{R}$. Moreover, the constant c_α appearing in Theorem 2 can be written down explicitly but we did not attempt to optimize the constant. The proof of Theorem 2 is along the same lines as

Theorem 1, but somewhat more involved. Deferring the proof to Section 3, let us move on to examples.

2.1. **Example: Curie-Weiss model at criticality.** The 'Curie-Weiss model of ferromagnetic interaction' at inverse temperature β and zero external field is given by the following Gibbs measure on $\{+1,-1\}^n$. For a typical configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in \{+1,-1\}^n$ the probability of $\boldsymbol{\sigma}$ is given by

$$\mu_\beta(\{\boldsymbol{\sigma}\}) := Z_\beta^{-1} \exp\left(\frac{\beta}{n}\sum_{i<j}\sigma_i\sigma_j\right)$$

where $Z_\beta = Z_\beta(n)$ is the normalizing constant. It is well known that the Curie-Weiss model shows a phase transition at $\beta_c = 1$. For $\beta < \beta_c$ the magnetization $m(\boldsymbol{\sigma}) := \frac{1}{n}\sum_{i=1}^n \sigma_i$ is concentrated at 0 but for $\beta > \beta_c$ the magnetization is concentrated on the set $\{-x^*, x^*\}$ where $x^* > 0$ is the largest solution of the equation $x = \tanh(\beta x)$. In fact using concentration inequalities for exchangeable pairs it was proved in [10] (Proposition 1.3) that for all $\beta \geq 0, h \in \mathbb{R}, n \geq 1, t \geq 0$ we have

$$\mathbb{P}\left(|m - \tanh(\beta m + h)| \geq \frac{\beta}{n} + \frac{t}{\sqrt{n}}\right) \leq 2\exp\left(-\frac{t^2}{4(1+\beta)}\right),$$

where h is the external field, which is zero in our case. Although a lot is known about this model (see Ellis [19] Section IV.4 for a survey), the above result – to the best of our knowledge – is the first rigorously proven concentration inequality that holds at all temperatures. (See also [14] for some related results.)

Incidentally, the above result shows that when $\beta < 1$, the magnetization is at most of order $n^{-1/2}$. It is known that at the critical temperature the magnetization $m(\boldsymbol{\sigma})$ shows a non Gaussian behavior and is of order $n^{-1/4}$. In fact, at $\beta = 1$ as $n \to \infty$, $n^{1/4}m(\boldsymbol{\sigma})$ converges to the probability distribution on $\mathbb{R}$ having density proportional to $\exp(-t^4/12)$. This limit theorem was first proved by Simon and Griffiths [23] and error bounds were obtained recently [13, 16]. The following concentration inequality, derived using Theorem 2, fills the gap in the tail bound at the critical point.

Proposition 4. *Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature $\beta = 1$. Then, for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies*

$$\mathbb{P}(n^{1/4}|m(\boldsymbol{\sigma})| \geq t) \leq 2e^{-ct^4}$$

where $c > 0$ is an absolute constant.

Here we may remark that such a concentration inequality probably cannot be obtained by application of standard off-the-shelf results (e.g. those surveyed in Ledoux [30], the famous results of Talagrand [38] or the recent breakthroughs of Boucheron, Lugosi and Massart [9]), because they generally give Gaussian or exponential tail bounds. There are several recent remarkable results giving tail bounds different from exponential and Gaussian. The papers [14, 20, 29] deal with tails between exponential and Gaussian and [1, 4] deal with sub-exponential tails. Also in [5, 21, 22] the authors deal with tails (possibly) larger than Gaussian. However, it seems that none of the techniques given in these references would lead to the result of Proposition 4.

It is possible to derive a similar tail bound using the asymptotic results of Martin-Löf [31] about the partition function $Z_\beta(n)$ (see also Bolthausen [7]). An application of their results

gives that

$$\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^n} e^{\frac{n}{2}m(\boldsymbol{\sigma})^2+n\theta m(\boldsymbol{\sigma})^4} \simeq \frac{2^{n+1}\Gamma(5/4)}{\sqrt{2\pi}}\left(\frac{12n}{1-12\theta}\right)^{1/4}$$

for $\theta < 1/12$ in the sense that the ratio of the two sides converges to one as n goes to infinity and from here the tail bound follows easily (without an explicit constant). However this approach depends on a precise estimate of the partition function (for example, large deviation estimates or finding the limiting free energy $\lim n^{-1}\log Z_\beta(n)$ are not enough) and this precise estimate is hard to prove. Our method, on the other hand, depends only on simple properties of the Gibbs measure and is not tied specifically to the Curie-Weiss model.

The idea used in the proof of Proposition 4 can be used to prove a tail inequality that holds for all $0 \le \beta \le 1$. We state the result below without proof. Note that the inequality gives the correct tail bound for all $0 \le \beta \le 1$.

Proposition 5. *Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at inverse temperature β where $0 \le \beta \le 1$. Then, for any $n \ge 1$ and $t \ge 0$ the magnetization satisfies*

$$\mathbb{P}(3(1-\beta)m(\boldsymbol{\sigma})^2 + \beta^3 m(\boldsymbol{\sigma})^4 \ge t) \le 2e^{-nt/160}.$$

It is possible to derive similar non-Gaussian tail inequalities for general Curie-Weiss models at the critical temperature. We briefly discuss the general case below. Let ρ be a symmetric probability measure on $\mathbb{R}$ with $\int x^2\, d\rho(x) = 1$ and $\int \exp(\beta x^2/2)\, d\rho(x) < \infty$ for all $\beta \ge 0$. The general Curie-Weiss model $\mathrm{CW}(\rho)$ at inverse temperature β is defined as the array of spin random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ with joint distribution

$$d\nu_n(\mathbf{x}) = Z_n^{-1}\exp\left(\frac{\beta}{2n}(x_1+x_2+\cdots+x_n)^2\right)\prod_{i=1}^n d\rho(x_i) \tag{3}$$

for $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ where

$$Z_n = \int \exp\left(\frac{\beta}{2n}(x_1+x_2+\cdots+x_n)^2\right)\prod_{i=1}^n d\rho(x_i)$$

is the normalizing constant. The magnetization $m(\mathbf{x})$ is defined as usual by $m(\mathbf{x}) = n^{-1}\sum_{i=1}^n x_i$. Here we will consider the case when ρ satisfies the following two conditions:

(A) ρ has compact support, that is, $\rho([-L, L]) = 1$ for some $L < \infty$.

(B) The equation $h'(s) = 0$ has a unique root at $s = 0$ where

$$h(s) := \frac{s^2}{2} - \log\int \exp(sx)\, d\rho(x) \text{ for } s \in \mathbb{R}.$$

The second condition says that $h(\cdot)$ has a unique global minima at $s = 0$ and $|h'(s)| > 0$ for $|s| > 0$. The behavior of this model is quite similar to the classical Curie-Weiss model and there is a phase transition at $\beta = 1$. For $\beta < 1$, $m(\mathbf{X})$ is concentrated around zero while for $\beta > 1, m(\mathbf{X})$ is bounded away from zero a.s. (see Ellis and Newman [17, 18]). We will prove the following concentration result.

Proposition 6. *Suppose $\mathbf{X} \sim \nu_n$ at the critical temperature $\beta = 1$ where ρ satisfies condition* (A) *and* (B). *Let k be such that $h^{(i)}(0) = 0$ for $0 \le i < 2k$ and $h^{(2k)}(0) \neq 0$, where*

$$h(s) := \frac{s^2}{2} - \log\int \exp(sx)\, d\rho(x) \text{ for } s \in \mathbb{R}$$

and $h^{(i)}$ is the i-th derivative of h. Then, $k > 1$ and for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies

$$\mathbb{P}(n^{1/2k}|m(\mathbf{X})| \geq t) \leq 2e^{-ct^{2k}}$$

where $c > 0$ is an absolute constant depending only on ρ.

Here we mention that in Ellis and Newman [17], convergence results were proved for the magnetization in $\mathrm{CW}(\rho)$ model under optimal condition on ρ. Under our assumption their result says that $n^{1/2k}m(\mathbf{X})$ converges weakly to a distribution having density proportional to $\exp(-\lambda x^{2k}/(2k)!)$ where $\lambda := h^{(2k)}(0)$. Hence the tail bound gives the correct convergence rate.

Let us now give a brief sketch of the proof of Proposition 4. Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature. We construct $\boldsymbol{\sigma}'$ by taking one step in the heat-bath Glauber dynamics: A coordinate I is chosen uniformly at random, and σ_I is replace by σ_I' drawn from the conditional distribution of the I-th coordinate given $\{\sigma_j : j \neq I\}$. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := \sum_{i=1}^{n} (\sigma_i - \sigma_i') = \sigma_I - \sigma_I'.$$

For each $i = 1, 2, \ldots, n$, define $m_i = m_i(\boldsymbol{\sigma}) = n^{-1} \sum_{j \neq i} \sigma_j$. An easy computaion gives that $\mathbb{E}(\sigma_i | \{\sigma_j, j \neq i\}) = \tanh(m_i)$ for all i and so we have

$$f(\boldsymbol{\sigma}) := \mathbb{E}(F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')|\boldsymbol{\sigma}) = m - \frac{1}{n}\sum_{i=1}^{n} \tanh(m_i) = \frac{m}{n} + \frac{1}{n}\sum_{i=1}^{n} g(m_i)$$

where $g(x) := x - \tanh(x)$. Note that $|m_i - m| \leq 1/n$, and hence $f(\boldsymbol{\sigma}) = m - \tanh m + O(1/n)$. A simple analytical argument using the fact that, for $x \approx 0$, $x - \tanh x = x^3/3 + O(x^5)$ then gives

$$\Delta(\boldsymbol{\sigma}) \leq \frac{6}{n}|f(\boldsymbol{\sigma})|^{2/3} + \frac{12}{n^{5/3}}$$

and using Corollary 3 with $\alpha = 2/3, B = 6/n$ and $C = 12/n^{5/3}$ we have

$$\mathbb{P}(|m - \tanh m| \geq t + n^{-1}) \leq \mathbb{P}(|f(\boldsymbol{\sigma})| \geq t) \leq 2e^{-cnt^{4/3}}$$

for all $t \geq 0$ for some constant $c > 0$. It is easy to see that this implies the result. The critical observation, of course, is that $x - \tanh(\beta x) = O(x^3)$ for $\beta = 1$, which is not true for $\beta \neq 1$.

2.2. **Example: Triangles in Erdős-Rényi graphs.** Consider the Erdős-Rényi random graph model $G(n,p)$ which is defined as follows. The vertex set is $[n] := \{1, 2, \ldots, n\}$ and each edge (i,j), $1 \leq i < j \leq n$ is present with probability p and not present with probability $1-p$ independently of each other. For any three distinct vertex $i < j < k$ in $[n]$ we say that the triple (i,j,k) forms a triangle in the graph $G(n,p)$ if all the three edges $(i,j), (j,k), (i,k)$ are present in $G(n,p)$ (see figure 1). Let T_n be the number of triangles in $G(n,p)$, that is

$$T_n := \sum_{1 \leq i < j < k \leq n} \mathbf{1}\{(i,j,k) \text{ forms a triangle in } G(n,p)\}.$$

Let us define the function $I(\cdot, \cdot)$ on $(0,1) \times (0,1)$ as

$$I(r,s) := r \log \frac{r}{s} + (1-r) \log \frac{1-r}{1-s}. \tag{4}$$

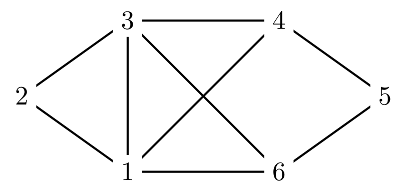

FIGURE 1. A graph on 6 vertices with 8 edges and 3 triangles. The triangles being $(1,2,3),(1,3,4)$ and $(1,3,6)$.

Note that $I(r,s)$ is the Kullback-Leibler divergence of the measure ν_s from ν_r and also the relative entropy of ν_r w.r.t. ν_s where ν_p is the Bernoulli(p) measure. We have the following result about the large deviation rate function for the number of triangles in $G(n,p)$.

Theorem 7. *Let T_n be the number of triangles in $G(n,p)$, where $p > p_0$ where $p_0 = 2/(2+e^{3/2}) \approx 0.31$. Then for any $r \in (p,1]$,*

$$\mathbb{P}\left(T_n \geq \binom{n}{3} r^3\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right). \tag{5}$$

Moreover, even if $p \leq p_0$, there exist p', p'' such that $p < p' \leq p'' < 1$ and the same result holds for all $r \in (p,p') \cup (p'',1]$. For all p and r in the above domains, we also have the more precise estimate

$$\mathbb{P}\left(\left|T_n - \binom{n}{3} r^3\right| \leq C(p,r) n^{5/2}\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right), \tag{6}$$

where $C(p,r)$ is a constant depending on p and r.

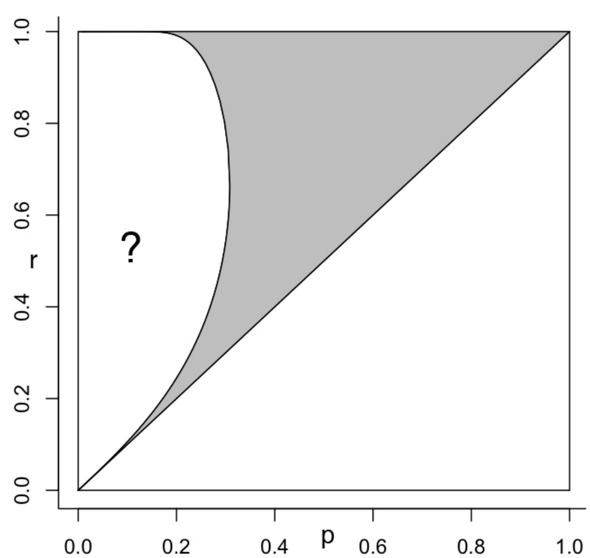

FIGURE 2. The set (colored in gray) of $(p,r), r \geq p$ for which we are able to show that the large deviation result holds.

The behavior of the upper tail of subgraph counts in $G(n,p)$ is a problem of great interest in the theory of random graphs (see [6, 25, 27, 28, 39], and references contained therein).

The best upper bounds to date were obtained by Kim and Vu [28] (triangles) and Janson, Oleszkiewicz, and Ruciński [26] (general subgraph counts). For triangles, the results of these papers essentially state that for a fixed $\epsilon > 0$,

$$\exp(-\Theta(n^2p^2\log(1/p))) \le \mathbb{P}(T_n \ge \mathbb{E}(T_n) + \epsilon n^3p^3) \le \exp(-\Theta(n^2p^2)).$$

Clearly, our result gives a lot more in the situations where it works (see Figure 2). The method of proof can be easily extended to prove similar results for general subgraph counts and are discussed in Subsection 2.3. However, there is an obvious incompleteness in Theorem 7 (and also for general subgraphs counts), namely, that it does not work for all (p, r).

In this context, we should mention that another paper on large deviations for subgraph counts by Bolthausen, Comets and Dembo [8] is in preparation. As of now, to the best of our knowledge, the authors of [8] have only looked at subgraphs that do not complete loops, like 2-stars. Another related article is the one by Döring and Eichelsbacher [15], who obtain moderate deviations for a class of graph-related objects, including triangles.

Unlike the previous two examples, Theorem 7 is far from being a direct consequence of any of our abstract results. Therefore, let us give a sketch of the proof, which involves a new idea.

The first step is standard: consider tilted measures. However, the appropriate tilted measure in this case leads to what is known as an 'exponential random graph', a little studied object in the rigorous literature. Exponential random graphs have become popular in the statistical physics and network communities in recent years (see the survey of Park and Newman [33]). The only rigorous work we are aware of is the recent paper of Bhamidi et. al. [3], who look at convergence rates of Markov chains that generate such graphs.

We will not go into the general definition or properties of exponential random graphs. Let us only define the model we need for our purpose.

Fix two numbers $\beta \ge 0$ and $h \in \mathbb{R}$. Let $\Omega = \{0,1\}^{\binom{n}{2}}$ be the space of all tuples like $\mathbf{x} = (x_{ij})_{1\le i<j\le n}$, where $x_{ij} \in \{0,1\}$ for each i, j. Let $\mathbf{X} = (X_{ij})_{1\le i<j\le n}$ be a random element of Ω following the probability measure proportional to $e^{H(\mathbf{x})}$, where H is the Hamiltonian

$$H(\mathbf{x}) = \frac{\beta}{n}\sum_{1\le i<j<k\le n} x_{ij}x_{jk}x_{ik} \;+\; h\sum_{1\le i<j\le n} x_{ij}.$$

Note that any element of Ω naturally defines an undirected graph on a set of n vertices. For each $\mathbf{x} \in \Omega$, let $T(\mathbf{x}) = \sum_{i<j<k} x_{ij}x_{jk}x_{ik}$ denote the number of triangles in the graph defined by $\mathbf{x}$, and let $E(\mathbf{x}) = \sum_{i<j} x_{ij}$ denote the number of edges. Then the above Hamiltonian is nothing but

$$\frac{\beta T(\mathbf{x})}{n} + hE(\mathbf{x}).$$

For notational convenience we will assume that $x_{ij} = x_{ji}$. Let $Z_n(\beta, h)$ be the corresponding partition function, that is

$$Z_n(\beta, h) = \sum_{\mathbf{x}\in\Omega} e^{H(\mathbf{x})}.$$

Note that $\beta = 0$ corresponds to the Erdős-Rényi random graph with $p = e^h/(1 + e^h)$. The following theorem 'solves' this model in a 'high temperature region'. Once this solution is known, the computation of the large deviation rate function is just one step away.

Theorem 8 (Free energy in high temperature regime)**.** *Suppose we have $\beta \ge 0$, $h \in \mathbb{R}$, and $Z_n(\beta, h)$ defined as above. Define a function $\varphi : [0,1] \to \mathbb{R}$ as*

$$\varphi(x) = \frac{e^{\beta x+h}}{1 + e^{\beta x+h}}.$$

Suppose β and h are such that the equation $u = \varphi(u)^2$ has a unique solution u^* in $[0,1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$. Then

$$\lim_{n\to\infty} \frac{\log Z_n(\beta,h)}{n^2} = -\frac{1}{2}I(\varphi(u^*),\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u^*)^3}{6},$$

where $I(\cdot,\cdot)$ is the function defined in (4). Moreover, there exists a constant $K(\beta,h)$ that depends only on β and h (and not on n) such that difference between $n^{-2}\log Z_n(\beta,h)$ and the limit is bounded by $K(\beta,h)n^{-1/2}$ for all n.

Incidentally, the above solution was obtained using physical heuristics by Park and Newman [34] in 2005. Here we mention that, in fact, the following result is always true.

Lemma 9. *For any $\beta \geq 0, h \in \mathbb{R}$ we have*

$$\begin{aligned}\liminf_{n\to\infty} \frac{\log Z_n(\beta,h)}{n^2} &\geq \sup_{r\in(0,1)} \left\{ -\frac{1}{2}I(r,\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta r^3}{6} \right\} \\ &= \sup_{u:\varphi(u)^2=u} \left\{ -\frac{1}{2}I(\varphi(u),\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u)^3}{6} \right\}.\end{aligned} \tag{7}$$

We will characterize the set of β, h for which the conditions in Theorem 8 hold in Lemma 12. First of all, note that the appearance of the function $\varphi(u)^2 - u$ is not magical. For each $i < j$, define

$$L_{ij} = \frac{1}{n}\sum_{k\notin\{i,j\}} X_{ik}X_{jk}.$$

This is the number of 'wedges' or 2-stars in the graph that have the edge ij as base. The key idea is to use Theorem 1 to show that these quantities approximately satisfy the following set of 'mean field equations':

$$L_{ij} \simeq \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \text{ for all } i<j. \tag{8}$$

(The idea of using Theorem 1 to prove mean field equations was initially developed in Section 3.4 of [10].) The following lemma makes this notion precise. Later, we will show that under the conditions of Theorem 8, this system has a unique solution.

Lemma 10 (Mean field equations)**.** *Let φ be defined as in Theorem 8. Then for any $1 \leq i < j \leq n$, we have*

$$\mathbb{P}\left(\sqrt{n}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk})\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{8(1+\beta)}\right)$$

for all $t \geq 8\beta/n$. In particular we have

$$\mathbb{E}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk})\right| \leq \frac{C(1+\beta)^{1/2}}{n^{1/2}} \tag{9}$$

where C is a universal constant.

In fact, one would expect that $L_{ij} \simeq u^*$ for all $i < j$, if the equation

$$\psi(u) := \varphi(u)^2 - u = 0 \tag{10}$$

has a unique solution u^* in $[0,1]$. The intuition behind is as follows. Define $L_{\max} = \max_{i,j} L_{ij}$ and $L_{\min} = \min_{i,j} L_{ij}$. It is easy to see that φ is an increasing function. Hence from the

mean-field equations (8) we have $L_{\max} \le \varphi(L_{\max})^2 + o(1)$ or $\psi(L_{\max}) \ge o(1)$. But $\psi(u) \ge 0$ iff $u \le u^*$. Hence $L_{\max} \le u^* + o(1)$. Similarly we have $L_{\min} \ge u^* - o(1)$ and thus all $L_{ij} \simeq u^*$. Lemma 11 formalizes this idea. Here we mention that one can easily check that equation (10) has at most three solutions. Moreover, $\psi(0) > 0 > \psi(1)$ implies that $\psi'(u^*) \le 0$ or $2\varphi(u^*)\varphi'(u^*) \le 1$ if u^* is the unique solution to (10).

Lemma 11. *Let u^* be the unique solution of the equation $u = \varphi(u)^2$. Assume that $2\varphi(u^*)\varphi'(u^*) < 1$. Then for each $1 \le i < j \le n$, we have*

$$\mathbb{E}\,|L_{ij} - u^*| \le \frac{K(\beta,h)}{n^{1/2}}$$

where $K(\beta,h)$ is a constant depending only on β,h. Moreover, if $2\varphi(u^*)\varphi'(u^*) = 1$ then we have

$$\mathbb{E}\,|L_{ij} - u^*| \le \frac{K(\beta,h)}{n^{1/6}} \text{ for all } 1 \le i < j \le n.$$

Now observe that the Hamiltonian $H(\mathbf{X})$ can be written as

$$H(\mathbf{X}) = \frac{\beta}{6} \sum_{1\le i<j\le n} X_{ij} L_{ij} + h \sum_{1\le i<j\le n} X_{ij}.$$

The idea then is the following: once we know that the conclusion of Lemma 11 holds, each L_{ij} in the above Hamiltonian can be replaced by u^*, which results in a model where the coordinates are independent. The resulting probability measure is presumably quite different from the original measure, but somehow the partition functions remain comparable.

The following lemma (Lemma 12) characterizes the region $S \in \mathbb{R} \times [0,\infty)$ such that the equation $u = \varphi(u)^2$ has a unique solution u^* in $[0,1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$ for $(h,\beta) \in S$ (see figure 3).

Let $h_0 = \log 2 - \frac{3}{2} < 0$. For $h < h_0$ there exist exactly two solutions $0 < a_* = a_*(h) < 1/2 < a^* = a^*(h) < \infty$ to the equation

$$\log x + \frac{1+x}{2x} + h = 0.$$

Define $a_*(h) = a^*(h) = 1/2$ for $h = h_0$ and

$$\beta_*(h) = \frac{(1+a_*)^3}{2a_*} \text{ and } \beta^*(h) = \frac{(1+a^*)^3}{2a^*} \tag{11}$$

for $h \le h_0$.

Lemma 12 (Characterization of high temperature regime)**.** *Let S be the set of pairs (h,β) for which the function $\psi(u) := \varphi(u)^2 - u$ has a unique root u^* in $[0,1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$ where $\varphi(u) := e^{\beta u + h}/(1 + e^{\beta u + h})$. Then we have*

$$S^c = \{(h,\beta) : h \le h_0 \text{ and } \beta_*(h) \le \beta \le \beta^*(h)\}$$

where β^*, β_* are as given in equation (11). In particular, $(h,\beta) \in S$ if $\beta \le (3/2)^3$ or $h > h_0$.

Remark. The point $h = h_0, \beta = \beta_0 := (3/2)^3$ is the critical point and the curve

$$\gamma(t) = \left(-\log t - \frac{1+t}{2t}, \frac{(1+t)^3}{2t}\right) \tag{12}$$

for $t > 0$ is the phase transition curve. It corresponds to $\psi(u^*) = 0$ and $2\psi(u^*)\psi'(u^*) = 1$. In fact, at the critical point (h_0, β_0) the function $\psi(u) = \varphi(u)^2 - u$ has a unique root of order

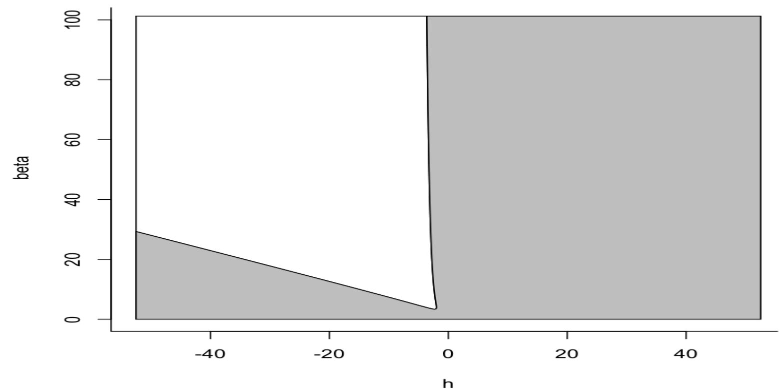

FIGURE 3. The set S (colored in gray) of (h, β) for which the conditions of Theorem 8 hold.

three at $u^* = 4/9$, *i.e.*, $\psi(u^*) = \psi'(u^*) = \psi''(u^*) = 0$ and $\psi'''(u^*) < 0$. The second part of lemma 11 shows that all the above conclusions (including the limiting free energy result) are true for the critical point but with an error rate of $n^{-1/6}$. Define the "energy" function

$$e(r) = \frac{1}{2} I(r, \varphi(0)) + \frac{1}{2}\log(1 - \varphi(0)) - \frac{\beta r^3}{6}$$

appearing in of the r.h.s. of equation (7). The "high temperature" regime corresponds to the case when $e(\cdot)$ has a unique minima and no local maxima or saddle point. The critical point corresponds to the case when $e(\cdot)$ has a non-quadratic global minima. The boundary corresponds to the case when $e(\cdot)$ has a unique minima and a saddle point. In the "low temperature" regime $e(\cdot)$ has two local minima. In fact, one can easily check that there is a one dimensional curve inside the set S^c, starting from the critical point, on which $e(\cdot)$ has two global minima and outside one global minima. Below we provide the solution on the boundary curve. Unfortunately, as of now, we don't have a rigorous solution in the "low temperature" regime.

For (h, β) on the phase transition boundary curve (excluding the critical point) the function $\psi(\cdot)$ has two roots and one of them, say v^*, is an inflection point. Let u^* be the other root. Here we mention that u^* is a minima of $e(\cdot)$ while v^* is a saddle point of $e(\cdot)$. On the lower part of the boundary, which corresponds to $\{\gamma(t) : t < 1/2\}$, the inflection point $v^* = (1+t)^{-2}$ is larger than u^*, while on the upper part of the boundary corresponding to $\{\gamma(t) : t > 1/2\}$, the inflection point $v^* = (1+t)^{-2}$ is smaller than u^*. The following lemma "solves" the model at the boundary point $\gamma(t)$ (see eqn. 12).

Lemma 13. *Let $\gamma(\cdot), u^*, v^*$ be as above and $(h, \beta) = \gamma(t)$ for some $t \neq 1/2$. Then, for each $1 \leq i < j \leq n$, we have*

$$\mathbb{E}(|L_{ij} - u^*|) \leq \frac{K(\beta, h)}{n^{1/2}} \tag{13}$$

for some constant $K(\beta, h)$ depending on β, h. Moreover, we have

$$\frac{\log Z_n(\beta, h)}{n^2} = -\frac{1}{2} I(\varphi(u^*), \varphi(0)) - \frac{1}{2}\log(1 - \varphi(0)) + \frac{\beta \varphi(u^*)^3}{6} + O(n^{-1/2})$$

and

$$\mathbb{P}\left(\left|T_n(\mathbf{Y}) - \binom{n}{3}\varphi(u^*)^3\right| \le C(\beta,h)n^{5/2}\right)$$
$$= \exp\left(-\frac{n^2 I(\varphi(u^*),\varphi(0))}{2}(1+O(n^{-1/2}))\right), \tag{14}$$

where $\mathbf{Y} = ((Y_{ij}))_{i<j}$ *follows* $G(n,\varphi(0))$ *and the constant appearing in* $O(\cdot)$ *and* $C(\beta,h)$ *depend only on* β,h.

In the next subsection we will briefly discuss about the results for general subgraph counts that can be proved using similar ideas.

2.3. **Example: General subgraph counts.** Let $F=(V(F),E(F))$ be a fixed finite graph on $\boldsymbol{v}_F := |V(F)|$ many vertices with $\boldsymbol{e}_F := |E(F)|$ many edges. Without loss of generality we will assume that $V(F) = [\boldsymbol{v}_F] := \{1,2,\ldots,\boldsymbol{v}_F\}$. Let $\alpha_F = |\mathrm{Aut}(F)|$ be the number of graph automorphism of the graph F. Let N_n be the number of copies of F, not necessarily induced, in the Erdős-Rényi random graph $G(n,p)$ (so the number of 2-stars in a triangle will be three). We have the following result about the large deviation rate function for the random variable N_n.

Theorem 14. *Let* N_n *be the number of copies of* F *in* $G(n,p)$, *where*

$$p > p_0 := \frac{\boldsymbol{e}_F - 1}{\boldsymbol{e}_F - 1 + \exp\left(\frac{\boldsymbol{e}_F}{\boldsymbol{e}_F - 1}\right)}.$$

Then for any $r \in (p,1]$,

$$\mathbb{P}\left(N_n \ge \frac{\boldsymbol{v}_F!}{\alpha_F}\binom{n}{\boldsymbol{v}_F} r^{\boldsymbol{e}_F}\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right). \tag{15}$$

Moreover, even if $p \le p_0$, *there exist* p', p'' *such that* $p < p' \le p'' < 1$ *and the same result holds for all* $r \in (p,p') \cup (p'',1]$. *For all* p *and* r *in the above domains, we also have the more precise estimate*

$$\mathbb{P}\left(\left|N_n - \frac{\boldsymbol{v}_F!}{\alpha_F}\binom{n}{\boldsymbol{v}_F} r^{\boldsymbol{e}_F}\right| \le C(p,r)n^{\boldsymbol{v}_F - 1/2}\right)$$
$$= \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right),$$

where $C(p,r)$ *is a constant depending on* p *and* r.

Note that p_0 as a function of $\boldsymbol{e}_F$ is increasing and converges to 1 as number of edges goes to infinity (see Figure 4). So there is an obvious gap in the large deviation result, namely the proof does not work when $r \ge p,\ p \le p_0$ and the gap becomes larger as the number of edges in F increases. Note that $p_0 \to 1$ as $\boldsymbol{e}_F \to \infty$.

The proof of Theorem 14 uses the same arguments that were used in the triangle case. Here the tilted measure leads to an exponential random graph model where the Hamiltonian depends on number of copies of F in the random graph. Let $\beta \ge 0, h \in \mathbb{R}$ be two fixed numbers. As before we will identify elements of $\Omega := \{0,1\}^{\binom{n}{2}}$ with undirected graphs on a set of n vertices. For each $\mathbf{x} \in \Omega$, let $N(\mathbf{x})$ denote the number of copies of F in the graph defined by $\mathbf{x}$, and let $E(\mathbf{x}) = \sum_{i<j} x_{ij}$ denote the number of edges. Let $\mathbf{X} = (X_{ij})_{1\le i<j\le n}$ be

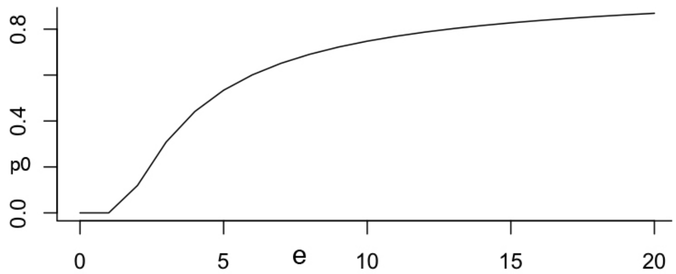

FIGURE 4. The curve p_0 vs. $\boldsymbol{e}_F$ where for a graph F with $\boldsymbol{e}_F$ many edges our large deviation result holds when $p > p_0$.

a random element of Ω following the probability measure proportional to $e^{H(\mathbf{x})}$, where H is the Hamiltonian

$$H(\mathbf{x}) = \frac{\beta}{(n-2)_{\boldsymbol{v}_F-2}} N(\mathbf{x}) + hE(\mathbf{x}).$$

where $(n)_m = \frac{n!}{(n-m)!}$. Recall that $\boldsymbol{v}_F$ is the number of vertices in the graph F. The scaling was done to make the two summands comparable. Also we used $(n-2)_{\boldsymbol{v}_F-2}$ instead of $n^{\boldsymbol{v}_F}$ to make calculations simpler. Let $Z_n(\beta, h)$ be the partition function. Note that $N(\mathbf{x})$ can be written as

$$N(\mathbf{x}) = \frac{1}{\alpha_F} \sum_{\substack{1 \le t_1, t_2, \ldots, t_{\boldsymbol{v}_F} \le n, \\ t_i \neq t_j \text{ for } i \neq j}} \prod_{(i,j) \in E(F)} x_{t_i t_j}. \tag{16}$$

For $\mathbf{x} \in \Omega, 1 \le i < j \le n$, define $\mathbf{x}^1_{(i,j)}$ as the element of Ω which is same as $\mathbf{x}$ in every coordinate except for the (i,j)-th coordinate where the value is 1. Similarly define $\mathbf{x}^0_{(i,j)}$. For $i < j$, define the random variable

$$L_{ij} := \frac{N(\mathbf{X}^1_{(i,j)}) - N(\mathbf{X}^0_{(i,j)})}{(n-2)_{\boldsymbol{v}_F-2}}.$$

The main idea is as in the triangle case. We show that L_{ij}'s satisfy a system of "mean-field equations" similar to (8) which has a unique solution under the condition of Theorem 15. In fact, we will show that L_{ij} "$\approx$" u^* for all $i < j$ and $E(\mathbf{X})$ "$\approx$" $\binom{n}{2}\varphi(u^*)$ under the condition of Theorem 15. Now note that we can write the hamiltonian as

$$H(\mathbf{X}) = \frac{\beta}{\boldsymbol{e}_F} \sum_{i<j} X_{ij} L_{ij} + h \sum_{i<j} X_{ij}$$

which is approximately equal to $h^* E(\mathbf{X})$ where $h^* = h + \beta u^*/\boldsymbol{e}_F$. Now the remaining is a calculus exercise.

So the first step in proving the large deviation bound is the following theorem, which gives the limiting free energy in the "high temperature" regime. Note the similarity with the triangle case.

Theorem 15. *Suppose we have $\beta \ge 0$, $h \in \mathbb{R}$, and $Z_n(\beta, h)$ defined as above. Define a function $\varphi : [0,1] \to \mathbb{R}$ as*

$$\varphi(x) = \frac{e^{\beta x + h}}{1 + e^{\beta x + h}}.$$

Suppose β and h are such that the equation $\alpha_F u = 2\boldsymbol{e}_F\varphi(u)^{\boldsymbol{e}_F-1}$ has a unique solution u^* in $[0,1]$ and $2\boldsymbol{e}_F(\boldsymbol{e}_F-1)\varphi(u^*)^{\boldsymbol{e}_F-2}\varphi'(u^*) < \alpha_F$. Then

$$\lim_{n\to\infty}\frac{\log Z_n(\beta,h)}{n^2} = -\frac{1}{2}I(\varphi(u^*),\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u^*)^{\boldsymbol{e}_F}}{\alpha_F},$$

where $I(\cdot,\cdot)$ is the function defined in (4). Moreover, there exists a constant $K(\beta,h)$ that depends only on β and h (and not on n) such that difference between $n^{-2}\log Z_n(\beta,h)$ and the limit is bounded by $K(\beta,h)n^{-1/2}$ for all n.

Here also we can identify the region where the conditions in Theorem 15 hold. Let

$$h_0 = \log(\boldsymbol{e}_F-1) - \frac{\boldsymbol{e}_F}{\boldsymbol{e}_F-1}. \tag{17}$$

For $h < h_0$ there exist exactly two solutions $0 < a_* = a_*(h) < 1/2 < a^* = a^*(h) < \infty$ of the equation

$$\log x + \frac{1+x}{(\boldsymbol{e}_F-1)x} + h = 0$$

Define $a_*(h) = a^*(h) = 1/(\boldsymbol{e}_F-1)$ for $h = h_0$ and

$$\beta_*(h) = \frac{\alpha_F(1+a_*)^{\boldsymbol{e}_F}}{2\boldsymbol{e}_F(\boldsymbol{e}_F-1)a_*} \text{ and } \beta^*(h) = \frac{\alpha_F(1+a^*)^{\boldsymbol{e}_F}}{2\boldsymbol{e}_F(\boldsymbol{e}_F-1)a^*} \tag{18}$$

for $h \le h_0$.

Lemma 16. *Let S be the set of pairs (h,β) for which the function*

$$\psi(u) := 2\boldsymbol{e}_F\varphi(u)^{\boldsymbol{e}_F-1} - \alpha_F u$$

has a unique root u^* in $[0,1]$ and $2\boldsymbol{e}_F(\boldsymbol{e}_F-1)\varphi(u^*)^{\boldsymbol{e}_F-2}\varphi'(u^*) < \alpha_F$ where $\varphi(u) := e^{\beta u+h}/(1+e^{\beta u+h})$. Then we have

$$S^c = \{(h,\beta) : h \le h_0 \text{ and } \beta_*(h) \le \beta \le \beta^*(h)\}$$

where h_0, β^*, β_* are as given in equations (17), (18). In particular, $(h,\beta)\in S$ if

$$\beta \le \frac{\alpha_F \boldsymbol{e}_F^{\boldsymbol{e}_F-1}}{2(\boldsymbol{e}_F-1)^{\boldsymbol{e}_F}} \text{ or } h > h_0.$$

In fact Lemma 16 identifies the critical point and the phase transition curve where the model goes from ordered phase to a disordered phase. But the results above does not say what happens at the boundary or in the low temperature regime. However note that the mean-field equations hold for all values of β and h.

2.4. **Example: Ising model on $\mathbb{Z}^d$.** Fix any $\beta \ge 0, h \in \mathbb{R}$ and an integer $d \ge 1$. Also fix $n \ge 2$. Let $\mathbb{B} = \{1,2,\ldots,n+1\}^d$ be a hypercube with $(n+1)^d$ many points in the d-dimensional hypercube lattice $\mathbb{Z}^d$. Let Ω be the graph obtained from $\mathbb{B}$ by identifying the opposite boundary points, *i.e.,* for $x = (x_1,x_2,\ldots,x_d), y = (y_1,y_2,\ldots,y_d) \in \mathbb{B}$ we have x is identified with y if $x_i - y_i \in \{-n,0,n\}$ for all i. This identification is known in the literature as periodic boundary condition. Note that Ω is the d-dimensional lattice torus with linear size n. We will write $x \sim y$ for $x,y\in\Omega$ if x,y are nearest neighbors in Ω. Also let us denote by N_x the set of nearest neighbors of x in Ω, *i.e.,* $N_x = \{y\in\Omega : y\sim x\}$.

Now consider the Gibbs measure on $\{+1,-1\}^\Omega$ given by the following Hamiltonian

$$H(\boldsymbol{\sigma}) := \beta\sum_{x\sim y, x,y\in\Omega}\sigma_x\sigma_y + h\sum_{x\in\Omega}\sigma_x$$

where $\boldsymbol{\sigma} = (\sigma_x)_{x\in\Omega}$ is a typical element of $\{+1,-1\}^\Omega$. So the probability of a configuration $\boldsymbol{\sigma} \in \{+1,-1\}^\Omega$ is

$$\mu_{\beta,h}(\{\boldsymbol{\sigma}\}) := Z_{\beta,h}^{-1}\exp\left(H(\boldsymbol{\sigma})\right) = Z_{\beta,h}^{-1}\exp\left(\beta\sum_{x\sim y, x,y\in\Omega}\sigma_x\sigma_y + h\sum_{x\in\Omega}\sigma_x\right) \tag{19}$$

where $Z_{\beta,h} = \sum_{\boldsymbol{\sigma}\in\{+1,-1\}^\Omega} e^{H(\boldsymbol{\sigma})}$ is the normalizing constant. Here σ_x is the spin of the magnetic particle at position x in the discrete torus Ω. This is the famous Ising model of ferromagnetism on the box $\mathbb{B}$ with periodic boundary condition at inverse temperature β and external field h.

The one-dimensional Ising model is probably the first statistical model of ferromagnetism to be proposed or analyzed [24]. The model exhibits no phase transition in one dimension. But for dimensions two and above the Ising ferromagnet undergoes a transition from an ordered to a disordered phase as β crosses a critical value. The two dimensional Ising model with no external field was first solved by Lars Onsager in a ground breaking paper [32], who also calculated the critical β as $\beta_c = \sinh^{-1}(1)$. For dimensions three and above the model is yet to be solved, and indeed, very few rigorous results are known.

In this subsection, we present some concentration inequalities for the Ising model that hold for all values of β. These 'temperature-free' relations are analogous to the mean field equations that we obtained for subgraph counts earlier.

The magnetization of the system, as a function of the configuration $\boldsymbol{\sigma}$, is defined as $m(\boldsymbol{\sigma}) := \frac{1}{|\Omega|}\sum_{x\in\Omega}\sigma_x$. For each integer $k \in \{1, 2, \ldots, 2d\}$, define a degree k polynomial function $r_k(\boldsymbol{\sigma})$ of a spin configuration $\boldsymbol{\sigma}$ as follows:

$$r_k(\boldsymbol{\sigma}) := \left(\binom{2d}{k}|\Omega|\right)^{-1}\sum_{x\in\Omega}\sum_{S\subseteq N_x, |S|=k}\sigma_S \tag{20}$$

where $\sigma_S = \prod_{x\in S}\sigma_x$ for any $S \subseteq \Omega$. In particular $r_k(\boldsymbol{\sigma})$ is the average of the product of spins of all possible k out of $2d$ neighbors. Note that $r_1(\boldsymbol{\sigma}) \equiv m(\boldsymbol{\sigma})$. We will show that when $h = 0$ and n is large, $m(\boldsymbol{\sigma})$ and $r_k(\boldsymbol{\sigma})$'s satisfy the following "mean-field relation" with high probability under the Gibbs measure:

$$(1-\theta_0(\beta))m(\boldsymbol{\sigma}) \approx \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma}). \tag{21}$$

These relations hold for all values of $\beta \geq 0$. Here θ_k's are explicit rational functions of $\tanh(2\beta)$ for $k = 0, 1, \ldots, d-1$, defined in equation (22) below. (Later we will prove in Proposition 19 that an external magnetic field h will add an extra linear term in the above relation (21).) The following Proposition makes this notion precise in terms of finite sample tail bound. It is a simple consequence of Theorem 1.

Theorem 17. *Suppose $\boldsymbol{\sigma}$ is drawn from the Gibbs measure $\mu_{\beta,0}$. Then, for any $\beta \geq 0, n \geq 1$ and $t \geq 0$ we have*

$$\mathbb{P}\left(\sqrt{|\Omega|}\left|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma})\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{4b(\beta)}\right)$$

where $m(\boldsymbol{\sigma}) := \frac{1}{|\Omega|}\sum_{x\in\Omega}\sigma_x$ *is the magnetization,* $r_k(\boldsymbol{\sigma})$ *is as given in* (20) *and for* $k = 0,1,\ldots,d-1$

$$\theta_k(\beta) = \frac{1}{4^d}\binom{2d}{2k+1}\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}}\tanh\left(\beta\sum_{i=1}^{2d}\sigma_i\right)\prod_{j=1}^{2k+1}\sigma_j$$

$$\text{and } b(\beta) = |1-\theta_0(\beta)| + \sum_{k=1}^{d-1}(2k+1)|\theta_k(\beta)|. \tag{22}$$

Moreover, we can explicitly write down $\theta_0(\beta)$ *as*

$$\theta_0(\beta) = \frac{1}{4^{d-1}}\sum_{k=1}^{d} k\binom{2d}{d+k}\tanh(2k\beta)$$

and for $d\geq 2$ *there exists* $\beta_1\in(0,\infty)$*, depending on* d*, such that* $1-\theta_0(\beta)>0$ *for* $\beta<\beta_1$ *and* $1-\theta_0(\beta)<0$ *for* $\beta>\beta_1$*.*

Here we may remark that for any fixed k, $\theta_k(\beta/2d)$ converges to the coefficient of x^{2k+1} in the power series expansion of $\tanh(\beta x)$ and $2d\beta_1(d)\downarrow 1$ as $d\to\infty$. For small values of d we can explicitly calculate the θ_k's. For instance, in $d=2$,

$$\theta_0(\beta) = \frac{1}{2}\left(\tanh(4\beta)+2\tanh(2\beta)\right),\ \theta_1(\beta) = \frac{1}{2}\left(\tanh(4\beta)-2\tanh(2\beta)\right).$$

For $d=3$,

$$\begin{aligned}
\theta_0(\beta) &= \frac{3}{16}\left(\tanh(6\beta)+4\tanh(4\beta)+5\tanh(2\beta)\right),\\
\theta_1(\beta) &= \frac{10}{16}\left(\tanh(6\beta)-3\tanh(2\beta)\right),\\
\theta_2(\beta) &= \frac{3}{16}\left(\tanh(6\beta)-4\tanh(4\beta)+5\tanh(2\beta)\right).
\end{aligned}$$

For $d=4$,

$$\begin{aligned}
\theta_0(\beta) &= \frac{1}{16}\left(\tanh(8\beta)+6\tanh(6\beta)+14\tanh(4\beta)+14\tanh(2\beta)\right),\\
\theta_1(\beta) &= \frac{7}{16}\left(\tanh(8\beta)+2\tanh(6\beta)-2\tanh(4\beta)-6\tanh(2\beta)\right),\\
\theta_2(\beta) &= \frac{7}{16}\left(\tanh(8\beta)-2\tanh(6\beta)-2\tanh(4\beta)+6\tanh(2\beta)\right),\\
\theta_3(\beta) &= \frac{1}{16}\left(\tanh(8\beta)-6\tanh(6\beta)+14\tanh(4\beta)-14\tanh(2\beta)\right).
\end{aligned}$$

Corollary 18. *For the Ising model on* Ω *at inverse temperature* β *with no external magnetic field for all* $t\geq 0$ *we have,*

(i) *if* $d=1$*,*

$$\mathbb{P}(|m(\boldsymbol{\sigma})|\geq t)\leq 2\exp\left(-\frac{1}{4}|\Omega|(1-\tanh(2\beta))t^2\right)$$

(ii) *if* $d=2$*,*

$$\mathbb{P}(|[(1-u)^2-u^3]m(\boldsymbol{\sigma})+u^3r_3(\boldsymbol{\sigma})|\geq t)\leq 2\exp\left(-\frac{|\Omega|t^2}{32}\right)$$

where $u = \tanh(2\beta)$ *and* $r_3(\boldsymbol{\sigma}) = \frac{1}{4|\Omega|}\sum^* \sigma_x\sigma_y\sigma_z$ *where the sum* $\sum^*$ *is over all* $x, y, z \in \Omega$ *such that* $|x-y| = 2, |z-y| = 2, |x-z| = 2$.

(iii) *if* $d = 3$,

$$\mathbb{P}(|g(u)m(\boldsymbol{\sigma}) + 5u^3(1+u^2)r_3(\boldsymbol{\sigma}) - 3u^5 r_5(\boldsymbol{\sigma})| \geq t) \leq 2\exp\left(-c|\Omega|t^2\right)$$

where c *is an absolute constant,* $g(u) = 1 - 3u + 4u^2 - 9u^3 + 3u^4 - 3u^5$, $u = \tanh(2\beta)$ *and* r_3, r_5 *are as defined in* (20).

Although we do not yet know the significance of the above relations, it seems somewhat striking that they are not affected by phase transitions. The exponential tail bounds show that many such relations can hold simultaneously. For completeness, we state below the corresponding result for nonzero external field.

Proposition 19. *Suppose* $\boldsymbol{\sigma}$ *is drawn from the Gibbs measure* $\mu_{\beta,h}$. *Let* $r_k(\boldsymbol{\sigma})$, $\theta_k(\beta)$, $b(\beta)$ *be as in proposition* (17). *Then, for any* $\beta \geq 0, h \in \mathbb{R}, n \geq 1$ *and* $t \geq 0$ *we have*

$$\mathbb{P}\left(|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - g(\boldsymbol{\sigma})| \geq t\right) \leq 2\exp\left(-\frac{|\Omega|t^2}{4b(\beta)(1+\tanh|h|)}\right) \tag{23}$$

where

$$g(\boldsymbol{\sigma}) := \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma}) + \tanh(h)\left(1 - \sum_{k=0}^{d-1}\theta_k(\beta)s_{2k+1}(\boldsymbol{\sigma})\right)$$

and

$$s_k(\boldsymbol{\sigma}) := \left(\binom{2d}{k}|\Omega|\right)^{-1}\sum_{x\in\Omega}\sum_{S\subseteq N_x, |S|=k}\sigma_{S\cup\{x\}}$$

is the average of products of spins over all k*-stars for* $k = 1, 2, \ldots, 2d$ *and* Ω *is the discrete torus in* $\mathbb{Z}^d$ *with* n^d *many points.*

3. Proofs

3.1. **Proof of Proposition 4.** Instead of proving Theorem 2 first, let us see how it is applied to prove the result for the Curie-Weiss model at critical temperature. The proof is simply an elaboration of the sketch given at the end of Subsection 2.1.

Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at critical temperature. We construct $\boldsymbol{\sigma}'$ by taking one step in the heat-bath Glauber dynamics: A coordinate I is chosen uniformly at random, and σ_I is replace by σ'_I drawn from the conditional distribution of the I-th coordinate given $\{\sigma_j : j \neq I\}$. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := \sum_{i=1}^n(\sigma_i - \sigma'_i) = \sigma_I - \sigma'_I.$$

For each $i = 1, 2, \ldots, n$, define $m_i = m_i(\boldsymbol{\sigma}) = n^{-1}\sum_{j\neq i}\sigma_j$. An easy computaion gives that $\mathbb{E}(\sigma_i|\{\sigma_j, j\neq i\}) = \tanh(m_i)$ for all i and so we have

$$f(\boldsymbol{\sigma}) := \mathbb{E}(F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')|\boldsymbol{\sigma}) = m - \frac{1}{n}\sum_{i=1}^n\tanh(m_i) = \frac{m}{n} + \frac{1}{n}\sum_{i=1}^n g(m_i)$$

where $g(x) := x - \tanh(x)$. By definition $m_i(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma}) = \sigma_i/n$ and $m_i(\boldsymbol{\sigma}') - m(\boldsymbol{\sigma}) = (\sigma_i + \sigma_I - \sigma'_I)/n$ for all i. Hence using Taylor expansion upto first degree and noting that

$|g'(x)| = \tanh^2(x) \le x^2$ we have

$$|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \le \frac{2}{n}|g'(m(\boldsymbol{\sigma}))| + \frac{2 + 5\max_{|x|\le 1}|g''(x)|}{n^2}$$
$$\le \frac{2}{n}m(\boldsymbol{\sigma})^2 + \frac{6}{n^2}.$$

Clearly $|F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')| \le 2$. Thus we have

$$\Delta(\boldsymbol{\sigma}) := \frac{1}{2}\,\mathbb{E}[|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \cdot |F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')| \mid \boldsymbol{\sigma}] \le \frac{2}{n}m(\boldsymbol{\sigma})^2 + \frac{6}{n^2}.$$

Now it is easy to verify that $|x|^3 \le 5|x - \tanh x|$ for all $|x| \le 1$. Note that this is the place where we need $\beta = 1$. For $\beta \ne 1$, the linear term dominates in $m - \tanh(\beta m)$. Hence it follows that

$$m(\boldsymbol{\sigma})^2 \le 5^{2/3}|m(\boldsymbol{\sigma}) - \tanh m(\boldsymbol{\sigma})|^{2/3} \le 3|f(\boldsymbol{\sigma})|^{2/3} + 3n^{-2/3}$$

where in the last line we used the fact that $|f(\boldsymbol{\sigma}) - (m - \tanh m)| \le 1/n$ and $5^{2/3} < 3$. Thus

$$\Delta(\boldsymbol{\sigma}) \le \frac{6}{n}|f(\boldsymbol{\sigma})|^{2/3} + \frac{12}{n^{5/3}}$$

and using Corollary 3 with $\alpha = 2/3, B = 6/n$ and $C = 12/n^{5/3}$ we have

$$\mathbb{P}(|m - \tanh m| \ge t + n^{-1}) \le \mathbb{P}(|f(\boldsymbol{\sigma})| \ge t) \le 2e^{-cnt^{4/3}}$$

for all $t \ge 0$ for some constant $c > 0$. This clearly implies that

$$\mathbb{P}(|m| \ge t) \le \mathbb{P}(|m - \tanh m| \ge t^3/5) \le 2e^{-cnt^4}$$

for all $t \ge 0$ and for some absolute constant $c > 0$. Thus we are done.

3.2. **Proof of Proposition 6.** The proof is along the lines of proof of proposition 4. Suppose $\mathbf{X}$ is drawn from the distribution ν_n. We construct $\mathbf{X}'$ as follows: a coordinate I is chosen uniformly at random, and X_I is replace by X_I' drawn from the conditional distribution of the I-th coordinate given $\{X_j : j \ne I\}$. Let

$$F(\mathbf{X}, \mathbf{X}') := \sum_{i=1}^{n}(X_i - X_i') = X_I - X_I'.$$

For each $i = 1, 2, \ldots, n$, define $m_i(\mathbf{X}) = n^{-1}\sum_{j\ne i} X_j$. An easy computaion gives that $\mathbb{E}(X_i|\{X_j, j \ne i\}) = g(m_i)$ for all $i = 1, 2, \ldots, n$ where $g(s) = \frac{d}{ds}(\log\int \exp(x^2/2n + sx)\,d\rho(x))$ for $s \in \mathbb{R}$. So we have

$$f(\mathbf{X}) := \mathbb{E}(F(\mathbf{X}, \mathbf{X}')|\mathbf{X}) = m(\mathbf{X}) - \frac{1}{n}\sum_{i=1}^{n} g(m_i(\mathbf{X})).$$

Define the function

$$h(s) = \frac{s^2}{2} - \log\int \exp(sx)\,d\rho(x) \text{ for } s \in \mathbb{R}. \tag{24}$$

Clearly h is an even function. Recall that k is an integer such that $h^{(i)}(0) = 0$ for $0 \le i < 2k$ and $h^{(2k)}(0) \ne 0$. We have $k \ge 2$ since $h''(0) = 1 - \int x^2\,d\rho(x) = 0$.

Now using the fact that $\rho([-L,L])=1$ it is easy to see that $|f(\mathbf{X})-h'(m(\mathbf{X}))|\le c/n$ for some constant c depending on L only. In the subsequent calculations c will always denote a constant depending only on L that may vary from line to line. Similarly we have

$$\begin{aligned}|f(\mathbf{X})-f(\mathbf{X}')| &\le \frac{|X_I-X_I'|}{n}\left(|1-g'(m(\mathbf{X}))|+\frac{c(1+\sup_{|x|\le L}|g''(x)|)}{n}\right)\\ &\le \frac{2L}{n}|h''(m(\mathbf{X}))|+\frac{c}{n^2}.\end{aligned}$$

Note that $|h''(s)|\le cs^{2k-2}$ for some constant c for all $s\ge 0$. This follows since $\lim_{s\to 0}h''(s)/s^{2k-2}$ exists and $h''(\cdot)$ is a bounded function. Also $\lim_{s\to 0}|h'(s)|/|s|^{2k-1}=|h^{(2k)}(0)|\neq 0$ and $|h'(s)|>0$ for $s>0$. So we have $|h'(s)|\ge c|s|^{2k-1}$ for some constant $c>0$ and all $|s|\le L$. From the above results we deduce that

$$\begin{aligned}|f(\mathbf{X})-f(\mathbf{X}')| \le \frac{c}{n}|(m(\mathbf{X}))|^{2k-2}+\frac{c}{n^2} &\le \frac{c}{n}|h'(m(\mathbf{X}))|^{\frac{2k-2}{2k-1}}+\frac{c}{n^2}\\ &\le \frac{c}{n}|f(\mathbf{X})|^{\frac{2k-2}{2k-1}}+\frac{c}{n^{2-1/(2k-1)}}.\end{aligned}$$

Now the rest of the proof follows exactly as for the classical Curie-Weiss model.

3.3. **Proof of Theorem 7.** First, let us state and prove a simple technical lemma.

Lemma 20. *Let $x_1,\ldots,x_k,y_1,\ldots,y_k$ be real numbers. Then*

$$\max_{1\le i\le n}\left|\frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}}-\frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}}\right|\le 2\max_{1\le i\le n}|x_i-y_i|.$$

and

$$\left|\log\sum_{i=1}^k e^{x_i}-\log\sum_{i=1}^k e^{y_i}\right|\le \max_{1\le i\le k}|x_i-y_i|.$$

Proof. Fix $1\le i\le k$. For $t\in[0,1]$, let

$$h(t)=\frac{e^{tx_i+(1-t)y_i}}{\sum_{j=1}^k e^{tx_j+(1-t)y_j}}.$$

Then

$$h'(t)=\left[(x_i-y_i)-\frac{\sum_{j=1}^k(x_j-y_j)e^{tx_j+(1-t)y_j}}{\sum_{j=1}^k e^{tx_j+(1-t)y_j}}\right]h(t).$$

This shows that $|h'(t)|\le 2\max_i|x_i-y_i|$ for all $t\in[0,1]$ and completes the proof of the first assertion. The second inequality is proved similarly. □

Proof of Lemma 10. Fix two numbers $1\le i<j\le n$. Given a configuration $\mathbf{X}$, construct another configuration $\mathbf{X}'$ as follows. Choose a point $k\in\{1,\ldots,n\}\backslash\{i,j\}$ uniformly at random, and replace the pair (X_{ik},X_{jk}) with (X_{ik}',X_{jk}') drawn from the conditional distribution given the rest of the edges. Let L_{ij}' be the revised value of L_{ij}. From the form of the Hamiltonian it is now easy to read off that for $x,y\in\{0,1\}$,

$$\begin{aligned}&\mathbb{P}(X_{ik}'=x,X_{jk}'=y\mid\mathbf{X})\\ &\propto\exp\left(\beta xL_{ik}+\beta yL_{jk}+hx+hy-\frac{\beta}{n}xX_{ij}X_{jk}-\frac{\beta}{n}yX_{ij}X_{ik}+\frac{\beta}{n}xyX_{ij}\right).\end{aligned}$$

An application of Lemma 20 shows that the terms having β/n as coefficient can be 'ignored' in the sense that for each $x, y \in \{0,1\}$,

$$\left|\mathbb{P}(X'_{ik} = x, X'_{jk} = y \mid \mathbf{X}) - \frac{e^{\beta x L_{ik} + \beta y L_{jk} + hx + hy}}{(1+e^{\beta L_{ik}+h})(1+e^{\beta L_{jk}+h})}\right| \le \frac{2\beta}{n}$$

In particular,

$$|\,\mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}) - \varphi(L_{ik})\varphi(L_{jk})| \le \frac{2\beta}{n}. \tag{25}$$

Now,

$$\begin{aligned}\mathbb{E}(L_{ij} - L'_{ij} \mid \mathbf{X}) &= \frac{1}{n(n-2)}\sum_{k\notin\{i,j\}} (X_{ik}X_{jk} - \mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}))\\ &= \frac{1}{n-2}L_{ij} - \frac{1}{n(n-2)}\sum_{k\notin\{i,j\}} \mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}). \end{aligned}\tag{26}$$

Let $F(\mathbf{X},\mathbf{X}') = (n-2)(L_{ij} - L'_{ij})$ and $f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X},\mathbf{X}') \mid \mathbf{X})$. Let

$$g(\mathbf{X}) = L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}).$$

From (25) and (26) it follows that

$$|f(\mathbf{X}) - g(\mathbf{X})| \le \frac{2\beta}{n}. \tag{27}$$

Since X' has the same distribution as X, the same bound holds for $|f(X') - g(X')|$ as well. Now clearly, $|F(X,X')| \le 1$. Again, $|g(X) - g(X')| \le 2/n$, and therefore

$$|f(X) - f(X')| \le \frac{4(1+\beta)}{n}.$$

Combining everything, and applying Theorem 1 with $B = 0$ and $C = 2(1+\beta)/n$, we get

$$\mathbb{P}(|f(\mathbf{X})| \ge t) \le 2\exp\left(-\frac{nt^2}{4(1+\beta)}\right)$$

for all $t \ge 0$. From (27) it follows that

$$\mathbb{P}(|g(\mathbf{X})| \ge t) \le \mathbb{P}(|f(\mathbf{X})| \ge t - 2\beta/n) \le 2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t \ge 8\beta/n$. This completes the proof of the tail bound. The bound on the mean absolute value is an easy consequence of the tail bound. □

Proof of Lemma 11. The proof is in two steps. In the first step we will get an error bound of order $n^{-1/2}\sqrt{\log n}$. In the second step we will improve it to $n^{-1/2}$. Define

$$\Delta = \max_{1\le i<j\le n}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}}\varphi(L_{ik})\varphi(L_{jk})\right|.$$

By Lemma 10 and union bound we have

$$\mathbb{P}\left(\Delta \ge t\right) \le n^2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t \geq 8\beta/n$. Intuitively the above equation says that Δ is of the order of $\sqrt{\log n/n}$, in fact we have $\mathbb{E}(\Delta^2) = O(\log n/n)$. Clearly φ is an increasing function. Hence we have

$$\varphi(L_{\min})^2 - \Delta \leq L_{\min} \leq L_{\max} \leq \varphi(L_{\max})^2 + \Delta$$

where $L_{\max} = \max_{1\leq i<j\leq n} L_{ij}$ and $L_{\min} = \min_{1\leq i<j\leq n} L_{ij}$.

Now assume that there exists a unique solution u^* of the equation $\varphi(u)^2 = u$ with $2\varphi(u^*)\varphi'(u^*) < 1$. For ease of notations, define the function $\psi(u) = \varphi(u)^2 - u$. We have $\psi(0) > 0 > \psi(1)$, u^* is the unique solution to $\psi(u) = 0$ and $\psi'(u^*) < 0$. It is easy to see that $\psi'(u) = 0$ has at most three solution ($\psi'(u) = 2\beta\varphi(u)^2(1-\varphi(u)) - 1$ is a third degree polynomial in $\varphi(u)$ and φ is a strictly increasing function).

Hence there exist positive real numbers ε, δ such that $|\psi(u)| > \varepsilon$ if $|u - u^*| > \delta$. Note that $\psi(u) > 0$ if $u < u^*$ and $\psi(u) < 0$ is $u > u^*$. Decreasing ε, δ without loss of generality we can assume that

$$\inf_{0<|u-u^*|\leq\delta} \left[\frac{u-u^*}{-\psi(u)}\right] = c > 0. \tag{28}$$

This is possible because $\psi'(u^*) < 0$. Note that $\psi(L_{\max}) \geq -\Delta$ and $\psi(L_{\min}) \leq \Delta$. Thus we have

$$u^* - \delta \leq L_{\min} \leq L_{\max} \leq u^* + \delta$$

when $\Delta < \varepsilon$. Using (28), $u^* \leq L_{\max} \leq u^* + \delta$ implies that $|L_{\max} - u^*| \leq c\Delta$ and $u^* - \delta \leq L_{\min} \leq u^*$ implies that $|L_{\min} - u^*| \leq c\Delta$. Thus, when $\Delta < \varepsilon$, we have $|L_{\max} - u^*| \leq c\Delta$ and $|L_{\min} - u^*| \leq c\Delta$ and in particular, $|L_{ij} - u^*| \leq c\Delta$ for all $i < j$. So we can bound the L^2 distance of L_{ij} from u^* by

$$\mathbb{E}(L_{ij} - u^*)^2 \leq c^2\,\mathbb{E}(\Delta^2) + \mathbb{P}(\Delta \geq \varepsilon) \leq K(\beta, h)\frac{\log n}{n}$$

for all $i < j$.

Now let us move to the second step. Recall from (9) that

$$\mathbb{E}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk})\right| \leq \frac{C(1+\beta)^{1/2}}{n^{1/2}} \tag{29}$$

for all $i < j$. Let $D_{ij} = L_{ij} - u^*$. Using Taylor expansion around u^* upto degree one we have

$$\begin{aligned}
\varphi(L_{ik})\varphi(L_{jk}) - \varphi(u^*)^2 &= \varphi(u^*)(\varphi(L_{ik}) - \varphi(u^*)) + \varphi(u^*)(\varphi(L_{jk}) - \varphi(u^*)) \\
&\quad + (\varphi(L_{ik}) - \varphi(u^*))(\varphi(L_{jk}) - \varphi(u^*)) \\
&= \varphi(u^*)\varphi'(u^*)(D_{ik} + D_{jk}) + R_{ijk}
\end{aligned}$$

where $\mathbb{E}(|R_{ijk}|) \leq C\,\mathbb{E}(D_{ij}^2) \leq Cn^{-1}\log n$ for some constant C depending only on β, h. Thus

$$\begin{aligned}
\mathbb{E}&\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) - D_{ij} + \frac{\varphi(u^*)\varphi'(u^*)}{n}\sum_{k\notin\{i,j\}}(D_{ik} + D_{jk})\right| \\
&\leq \frac{2u^*}{n} + \frac{1}{n}\sum_{k\notin\{i,j\}} \mathbb{E}\,|R_{ijk}| \leq \frac{C\log n}{n}.
\end{aligned} \tag{30}$$

Here we used the fact that $u^* = \varphi(u^*)^2$. Combining (29) and (30) we have

$$\mathbb{E}\left|D_{ij} - \frac{\varphi(u^*)\varphi'(u^*)}{n}\sum_{k\notin\{i,j\}}(D_{ik} + D_{jk})\right| \leq \frac{C}{\sqrt{n}}$$

for all $i < j$. By symmetry, $\mathbb{E}\,|D_{ij}|$ is the same for all i, j. Thus finally we have

$$\mathbb{E}\,|L_{ij} - u^*| = \mathbb{E}\,|D_{ij}| \le \frac{1}{1 - 2\varphi(u^*)\varphi'(u^*)} \cdot \frac{C}{\sqrt{n}} = \frac{K(\beta, h)}{\sqrt{n}}$$

where $K(\beta, h)$ is a constant depending on β, h.

When $\psi(u) = 0$ has a unique solution at $u = u^*$ with $2\psi(u^*)\psi'(u^*) = 1$, which happens at the critical point $\beta = (3/2)^3, h = \log 2 - 3/2$, instead of equation (28) we have

$$\inf_{0<|u-u^*|\le\delta} \left[\frac{(u-u^*)^3}{-\psi(u)}\right] = c > 0$$

since $\psi(u^*) = \psi'(u^*) = \psi''(u^*) = 0$ and $\psi'''(u^*) < 0$. Then using a similar idea as above one can easily show that

$$\mathbb{E}\,|L_{ij} - u^*| \le K(\beta, h) n^{-1/6}$$

for some constant K depending on β, h. This completes the proof of the Lemma. □

Remark. The proof becomes lot easier if we have

$$c := \varphi(1) \cdot \sup_{0\le x\le 1} \frac{|\varphi(x) - \varphi(u^*)|}{|x - u^*|} < \frac{1}{2}. \tag{31}$$

This is because, by the triangle inequality we have

$$\begin{aligned}\sum_{i<j} |L_{ij} - u^*| &\le \sum_{i<j} \left| L_{ij} - \frac{1}{n} \sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \right| \\ &\quad + \sum_{i<j} \left(\frac{1}{n} \sum_{k\notin\{i,j\}} \left|\varphi(L_{ik})\varphi(L_{jk}) - u^*\right| + \frac{2u^*}{n} \right).\end{aligned} \tag{32}$$

Now recall that condition (31) says that $\varphi(1)|\varphi(x) - \varphi(u^*)| \le c|x - u^*|$ for all $x \in [0, 1]$. Moreover $L_{ij} \in [0, 1]$ for all i, j, and $u^* = \varphi(u^*)^2$. Thus,

$$\left|\varphi(L_{ik})\varphi(L_{jk}) - u^*\right| \le c|L_{ik} - u^*| + c|L_{jk} - u^*|.$$

Combining everything we get

$$\sum_{i<j} |L_{ij} - u^*| \le \frac{\sum_{i<j} \left| L_{ij} - \frac{1}{n} \sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \right| + nu^*}{1 - 2c}.$$

Taking expectation on both sides, and applying Lemma 10, we get

$$\sum_{i<j} \mathbb{E}\,|L_{ij} - u^*| \le \frac{C(1+\beta)n^{3/2}}{1 - 2c}.$$

And this gives the required result. In fact using basic calculus results one can easily check that condition (31) is satisfied when $h \ge 0$ or $\beta \le 2$.

Now we will prove that in the exponential random graph model, the number of edges and number of triangles also satisfy certain 'mean-field' relations.

Lemma 21. *Recall that $E(\mathbf{x})$ and $T(\mathbf{x})$ denote the number of edges and number of triangles in the graph defined by the edge configuration $\mathbf{x} \in \Omega$. If $\mathbf{X}$ is drawn from the Gibbs' measure*

in Theorem 8, we have the bound

$$\mathbb{E}\Big|E(\mathbf{X}) - \sum_{i<j}\varphi(L_{ij})\Big| \le C(1+\beta)^{1/2}n$$
$$\mathbb{E}\Big|\frac{T(\mathbf{X})}{n} - \frac{1}{3}\sum_{i<j}L_{ij}\varphi(L_{ij})\Big| \le C(1+\beta)^{1/2}n$$

where and C is a universal constant.

Proof. It is not difficult to see that

$$\mathbb{E}(X_{ij} \mid (X_{kl})_{(k,l)\neq(i,j)}) = \varphi(L_{ij}).$$

Let us create $\mathbf{X}'$ by choosing $1 \le i < j \le n$ uniformly at random and replacing X_{ij} with X'_{ij} drawn from the conditional distribution of X_{ij} given $(X_{kl})_{(k,l)\neq(i,j)}$. Let $F(\mathbf{X},\mathbf{X}') = \binom{n}{2}(X_{ij} - X'_{ij})$. Then

$$f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X},\mathbf{X}')|\mathbf{X}) = \sum_{k<l}(X_{kl} - \varphi(L_{kl})) = E(\mathbf{X}) - \sum_{k<l}\varphi(L_{kl}).$$

Now $|F(\mathbf{X},\mathbf{X}')| \le \binom{n}{2}$ and $|f(\mathbf{X}) - f(\mathbf{X}')| \le 1+\beta$. Here we used the fact that $|\varphi'(x)| \le \beta/4$. Combining the above result and Theorem 1 with $B = 0, C = \frac{1}{2}(1+\beta)\binom{n}{2}$, we get the required bound.

Similarly, if we define $F(\mathbf{X},\mathbf{X}') = \binom{n}{2}(X_{ij}L_{ij} - X'_{ij}L_{ij})$. Then

$$\begin{aligned} f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X},\mathbf{X}')|\mathbf{X}) &= \sum_{k<l}(X_{kl}L_{kl} - \varphi(L_{kl})L_{kl}) \\ &= \frac{3}{n}T(\mathbf{X}) - \sum_{k<l}\varphi(L_{kl})L_{kl}. \end{aligned}$$

Again, $|F(\mathbf{X},\mathbf{X}')| \le \binom{n}{2}$ and $|f(\mathbf{X}) - f(\mathbf{X}')| \le C(1+\beta)$. The bound follows easily as before. □

The following result is an easy corollary of Lemma 11 and Lemma 21.

Corollary 22. *Suppose the conditions of Theorem 8 are satisfied. Then we have*

$$\mathbb{E}\left|E(\mathbf{X}) - \frac{n^2\varphi(u^*)}{2}\right| \le Cn^{3/2} \text{ and } \mathbb{E}\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le Cn^{3/2}$$

where C is a constant depending only on β, h.

Lemma 23. *Suppose the conditions of Theorem 8 are satisfied. Let T_n be the number of triangles in the Erdős-Rényi graph $G(n,\varphi(0))$. Then there is a constant $K(\beta,h)$ depending only on β and h such that for all n*

$$\left|\frac{\log \mathbb{P}(|T_n - \binom{n}{3}\varphi(u^*)^3| \le K(\beta,h)n^{5/2})}{n^2} - \frac{-I(\varphi(u^*),\varphi(0))}{2}\right| \le \frac{K(\beta,h)}{\sqrt{n}}.$$

Proof. Let X be drawn from the Gibbs' measure in Theorem 8 with parameters β, h. From corollary 22 we see that there exists a constant $K(\beta,h)$ such that (for all n)

$$\mathbb{P}\left(\left|E(X) - \frac{n^2\varphi(u^*)}{2}\right| \le K(\beta,h)n^{3/2}\right) \ge \frac{3}{4}$$

and

$$\mathbb{P}\left(\left|\frac{T(X)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le K(\beta,h)n^{3/2}\right) \ge \frac{3}{4}.$$

Now let

$$A = \left\{x \in \{0,1\}^n : \left|\frac{T(x)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le K(\beta,h)n^{3/2}\right\}$$

and

$$B = A \cap \left\{x \in \{0,1\}^n : \left|E(x) - \frac{n^2\varphi(u^*)}{2}\right| \le K(\beta,h)n^{3/2}\right\}.$$

Now suppose $Y = (Y_{ij})_{1\le i<j\le n}$ is a collection of i.i.d. random variables satisfying $\mathbb{P}(Y_{ij} = 1) = 1 - \mathbb{P}(Y_{ij} = 0) = \varphi(0)$ and $Z = (Z_{ij})_{1\le i<j\le n}$ is another collection of i.i.d. random variables with $\mathbb{P}(Z_{ij} = 1) = 1 - \mathbb{P}(Z_{ij} = 0) = \varphi(u^*)$. Without loss of generality we can assume that $K(\beta,h)$ was chosen large enough to ensure that (again, for all n) $\mathbb{P}(Z \in A) \ge 1/2$ and $\mathbb{P}(Z \in B) \ge 1/2$. Now, it follows directly from the definition of A and Lemma 20 that

$$\begin{aligned}
&\left|\log\sum_{x\in A} e^{hE(x)} - \log\sum_{x\in A} e^{\frac{\beta T(x)}{n}+hE(x)} + \frac{\beta n^2\varphi(u^*)^3}{6}\right| \\
&= \left|\log\sum_{x\in A} e^{hE(x)+\frac{\beta n^2\varphi(u^*)^3}{6}} - \log\sum_{x\in A} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \qquad (33)\\
&\le \beta\max_{x\in A}\left|\frac{T(x)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le \beta K(\beta,h)n^{3/2}.
\end{aligned}$$

Next, observe that

$$\begin{aligned}
&\left|\log\sum_{x\in A} e^{\frac{\beta T(x)}{n}+hE(x)} - \log\sum_{x\in \Omega} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \\
&= |\log\mathbb{P}(X \in A)| \le |\log(3/4)|. \qquad (34)
\end{aligned}$$

Similarly we have

$$\begin{aligned}
&\left|\log\sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)} - \log\sum_{x\in \Omega} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \\
&= |\log\mathbb{P}(X \in B)| \le |\log(1/2)| \qquad (35)
\end{aligned}$$

where we used the fact that $\mathbb{P}(X \in A\cap C) \ge \mathbb{P}(X \in A) + \mathbb{P}(X \in C) - 1$. Combining the last two inequalities, we get

$$\left|\log\sum_{x\in A} e^{\frac{\beta T(x)}{n}+hE(x)} - \log\sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \le \log(8/3). \qquad (36)$$

Next, note that by the definition of B and Lemma 20, we have that for any h',

$$\begin{aligned}
&\left|\log\sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)} - \frac{n^2(h-h')\varphi(u^*)}{2} - \frac{\beta n^2\varphi(u^*)^3}{6} - \log\sum_{x\in B} e^{h'E(x)}\right| \\
&\le \sup_{x\in B}\left|\frac{\beta T(x)}{n} + hE(x) - \frac{n^2(h-h')\varphi(u^*)}{2} - \frac{\beta n^2\varphi(u^*)^3}{6} - h'E(x)\right| \\
&\le (\beta + |h-h'|)K(\beta,h)n^{3/2}. \qquad (37)
\end{aligned}$$

Now choose $h' = \log \frac{\varphi(u^*)}{1-\varphi(u^*)}$. Then

$$\left|\log \sum_{x\in B} e^{h'E(x)} - \log \sum_{x\in\Omega} e^{h'E(x)}\right| = |\log \mathbb{P}(Z \in B)| \le \log 2. \tag{38}$$

Adding up (33), (36), (37), and (38), and using the triangle inequality, we get

$$\left|\log \sum_{x\in A} e^{hE(x)} - \frac{n^2(h-h')\varphi(u^*)}{2} - \log \sum_{x\in\Omega} e^{h'E(x)}\right| \le K'(\beta,h)n^{3/2} \tag{39}$$

where $K'(\beta,h)$ is a constant depending only on β,h. For any $s \in \mathbb{R}$, a trivial verification shows that

$$\log \sum_{x\in\Omega} e^{sE(x)} = \binom{n}{2}\log(1+e^s).$$

Again, note that $\log \mathbb{P}(Y \in A) = \log \sum_{x\in A} e^{hE(x)} - \log \sum_{x\in\Omega} e^{hE(x)}$. Therefore it follows from inequality (39) that

$$\left|\frac{\log \mathbb{P}(Y\in A)}{n^2} - \frac{(h-h')\varphi(u^*) + \log(1+e^{h'}) - \log(1+e^h)}{2}\right| \le \frac{K'(\beta,h)}{\sqrt{n}}.$$

Now $h = \log \frac{\varphi(0)}{1-\varphi(0)}$ and $h' = \log \frac{\varphi(u^*)}{1-\varphi(u^*)}$. Also, $\log(1+e^h) = -\log(1-\varphi(0))$ and $\log(1+e^{h'}) = -\log(1-\varphi(u^*))$. Substituting these in the above expression, we get

$$\left|\frac{\log \mathbb{P}(Y\in A)}{n^2} - \frac{-I(\varphi(u^*),\varphi(0))}{2}\right| \le \frac{K'(\beta,h)}{\sqrt{n}}.$$

This completes the proof of the Lemma. □

We are now ready to finish the proof of Theorem 8.

Proof of Theorem 8. Note that by adding the terms in (35), (37), and (38) from the proof of Lemma 23, and applying the triangle inequality, we get

$$\left|\frac{\log Z_n(\beta,h)}{n^2} - \frac{(h-h')\varphi(u)}{2} - \frac{\beta\varphi(u)^3}{6} - \frac{1}{2}\log(1+e^{h'})\right| \le \frac{K(\beta,h)}{\sqrt{n}}.$$

This can be rewritten as

$$\left|\frac{\log Z_n(\beta,h)}{n^2} + \frac{I(\varphi(u),\varphi(0)) + \log(1-\varphi(0))}{2} - \frac{\beta\varphi(u)^3}{6}\right| \le \frac{K(\beta,h)}{\sqrt{n}}.$$

This completes the proof of Theorem 8. □

Note that the proof of Theorem 8 contains a proof for the lower bound in the general case. We provide the proof below for completeness.

Proof of Lemma 9. Fix any $r \in (0,1)$. Define the set B_r as

$$B_r = \left\{x \in \{0,1\}^n : \left|\frac{T(x)}{n} - \frac{n^2r^3}{6}\right| \le K(r)n^{3/2}, \left|E(x) - \frac{n^2r}{2}\right| \le K(r)n^{3/2}\right\}$$

where $K(r)$ is chosen in such a way that $\mathbb{P}(Z \in B_r) \ge 1/2$ where $Z = ((Z_{ij}))_{i<j}$ and Z_{ij}'s are i.i.d. Bernoulli(r). From the proof of Lemma 23 it is easy to see that

$$\left|\log \sum_{x\in B_r} e^{\frac{\beta T(x)}{n} + hE(x)} - \frac{n^2}{2}\left((h-h')r + \frac{\beta r^3}{3} + \log(1+e^{h'})\right)\right| \le K'n^{3/2}$$

where $h' = \log \frac{r}{1-r}$ and K' is a constant depending on β, h, r. Simplifying we have

$$\begin{aligned}\frac{2}{n^2}\log Z_n(\beta,h) &\geq \frac{2}{n^2}\log\sum_{x\in B_r} e^{\frac{\beta T(x)}{n}+hE(x)} \\ &\geq \frac{\beta r^3}{3}+\log(1-p)-I(r,p)-\frac{K'}{\sqrt{n}} \end{aligned}\tag{40}$$

for all r where $p = e^h/(1+e^h)$. Now taking limit as $n\to\infty$ and maximizing over r we have the first inequality (7). Given β, h, define the function

$$f(r) = \frac{\beta r^3}{3}+\log(1-p)-I(r,p)$$

where $p = e^h/(1+e^h)$. One can easily check that $f'(r) \gtreqless 0$ iff $\varphi(u)^2 - u \gtreqless 0$ for $u = r^2$. From this fact the second equality follows. ☐

Lemma 24. *Let T_n be the number of triangles in the Erdős-Rényi graph $G(n,\varphi(0))$. Then there is a constant $K(\beta,h)$ depending only on β and h such that for all n*

$$\frac{\log\mathbb{P}(T_n \geq \binom{n}{3}\varphi(u^*)^3)}{n^2} \leq \frac{-I(\varphi(u^*),\varphi(0))}{2}+\frac{K(\beta,h)}{\sqrt{n}}.$$

Proof. By Markov's inequality, we have

$$\frac{\log\mathbb{P}(T_n \geq \binom{n}{3}\varphi(u^*)^3)}{n^2} \leq -\frac{\beta}{n^3}\binom{n}{3}\varphi(u^*)^3+\frac{\mathbb{E}(e^{\beta T_n/n})}{n^2}.$$

From the last part of Theorem 8, it is easy to obtain an optimal upper bound of the second term on the right hand side, which finishes the proof of the Lemma. ☐

Proof of Theorem 7. Given p and r, if for all r' belonging to a small neighborhood of r there exist β and h satisfying the conditions of Theorem 8 such that $\varphi(0) = p$ and $\varphi(u^*) = r'$, then a combination of Lemma 23 and Lemma 24 implies the conclusion of Theorem 7. If $p \geq p_0 = 2/(2+e^{3/2})$, we can just choose $h \geq h_0 = -\log 2 - 3/2$ such that $p = e^h/(1+e^h)$ and conclude, from Theorem 8, Lemma 23 and Lemma 12, that the large deviations limit holds for any $\beta \geq 0$. Varying β between 0 and ∞, it is possible to get for any $r \geq p$ a β such that $\varphi(u^*) = r$.

For $p \leq p_0$, we again choose h such that $\varphi(0) = p$. Note that $h \leq h_0$. The large deviations limit should hold for any $r \geq p$ for which there exists $\beta > 0$ such that $r = \varphi(u^*) = \sqrt{u^*}$ and $(h,\beta)\in S$. It is not difficult to verify that given h, u^* is a continuously increasing function of β in the regime for which $(h,\beta)\in S$. Recall the settings of Lemma 12. Thus, the values of r that is allowed is in the set $(p,p_*)\cup(p^*,1]$, where p^*, p_* are the unique non-touching solutions to the equations

$$\sqrt{p^*} = \frac{e^{\beta_*(h)p^*+h}}{1+e^{\beta_*(h)p^*+h}},\ \sqrt{p_*} = \frac{e^{\beta^*(h)p_*+h}}{1+e^{\beta^*(h)p_*+h}}.$$

This completes the proof of Theorem 7. ☐

Finally, let us round up by proving Lemma 12.

Proof of Lemma 12. Fix $h\in\mathbb{R}$. Define the function

$$\psi(x;h,\beta) := \varphi(x;h,\beta)^2 - x$$

where

$$\varphi(x;h,\beta)=\frac{e^{\beta x+h}}{1+e^{\beta x+h}} \text{ for } x\in[0,1].$$

For simplicity, we will omit β,h in $\varphi(x;\beta,h)$ and $\psi(x;\beta,h)$ when there is no chance of confusion. Note that $\psi(0)>0>\psi(1)$. Hence the equation $\varphi(x;\beta,h)=0$ has at least one solution. Also we have $\psi'(x)=2\beta\varphi(x)^2(1-\varphi(x))-1$ and φ is strictly increasing. Hence the equation $\psi'(x)=0$ has at most three solutions. So either the function ψ is strictly decreasing or there exist two numbers $0<a<b<1$ such that ψ is strictly decreasing in $[0,a]\cup[b,1]$ and strictly increasing in $[a,b]$. From the above observations it is easy to see that the equation $\psi(x)=0$ has at most three solutions for any β,h. If $\psi(x)=0$ has exactly two solutions then $\psi'=0$ at one of the solution.

Let $u_*=u_*(h,\beta)$ and $u^*=u^*(h,\beta)$ be the smallest and largest solutions of $\psi(x;h,\beta)=0$ respectively. If $u_*=u^*$ we have a unique solution of $\psi(x)=0$. From the fact that $\frac{\partial}{\partial\beta}\psi(x;h,\beta)>0$ for all $x\in[0,1],\beta\geq 0,h\in\mathbb{R}$ we can deduce that given h, $u_*(h,\beta)$ and $u^*(h,\beta)$ are increasing functions of β. Note that u_* is left continuous and u^* is right continuous in β given h. Also note that given $h\in\mathbb{R}$, $u^*=u_*$ if $\beta>0$ is very small or very large. So we can define $\beta_*(h)$ and $\beta^*(h)$ such that for $\beta<\beta_*(h)$ and for $\beta>\beta^*(h)$ we have $u_*(h,\beta)=u^*(h,\beta)$. β_* is the largest and β^* is the smallest such number.

Therefore, we can deduce that at $\beta=\beta_*(h),\beta^*(h)$ the equation $\psi(x;h,\beta)=0$ has exactly two solutions. Thus we have two real numbers $x_*,x^*\in[0,1]$ such that

$$\varphi(x)^2=x \text{ and } 2\beta\varphi(x)^2(1-\varphi(x))=1$$

for $(x,\beta)=(x_*,\beta_*)$ or (x^*,β^*). Thus we have $2\beta x(1-\sqrt{x})=1$ and

$$h=\log\frac{\sqrt{x}}{1-\sqrt{x}}-\frac{1}{2(1-\sqrt{x})}$$

for $x=x_*,x^*$. Define $a_*=x_*^{-1/2}-1$ and $a^*=(x^*)^{-1/2}-1$. Note that $x=(1+a)^{-2},\beta=(1+a)^3/2a^2$ for $(x,a,\beta)=(x_*,a_*,\beta_*)$ or (x^*,a^*,β^*) and we have

$$h=-\log a-\frac{1+a}{2a} \tag{41}$$

for $a=a_*,a^*$. Now the function $g(x)=-\log x-(1+x)/2x$ is strictly increasing for $x\in(0,1/2]$ and strictly decreasing for $x\geq 1/2$. So equation (41) has no solution for $h\geq g(1/2)=\log 2-3/2=:h_0$. For $h<h_0$ equation (41) has exactly two solutions and for $h=h_0$ equation (41) has one solution. One can easily check that $\beta_*\leq\beta^*$ implies that $a_*\leq a^*$. Also from the fact that (41) has at most two solutions, we have that for $\beta\in(\beta_*,\beta^*)$ the equation $\psi(u)=0$ has exactly three solutions. □

3.4. **Proof of Lemma 13.** For simplicity we will prove the result only for the lower boundary part, that is, for $(h,\beta)=\gamma(t)$ with $t<1/2$. The proof for the upper boundary is similar. Fix $t<1/2$. Let us briefly recall the setup. The function $\psi(u)=\varphi(u)^2-u$ has two roots at $0<u^*<v^*<1$ and $\psi'(u_*)<0$ while $\psi'(v^*)=0,\psi''(v^*)<0$.

Define the function

$$f(r)=\frac{\beta r^3}{3}+\log(1-p)-I(r,p) \text{ for } r\in(0,1).$$

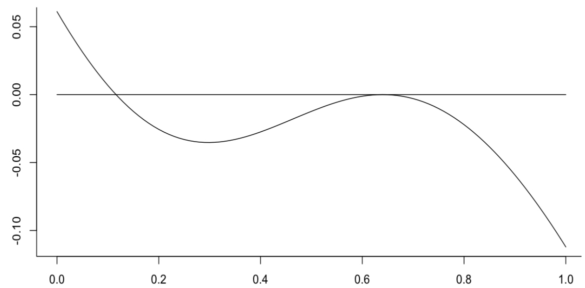

FIGURE 5. The function $\psi(\cdot)$ for $(h,\beta)=\gamma(1/4)$.

From the proof of Lemma 9 and the fact that $\psi'(u)<0$ for $u\in(u^*,v^*)$ it is easy to see that $f(\varphi(u^*))>f(\varphi(v^*))$ and

$$\frac{2}{n^2}\log Z_n(\beta,h)\ge f(\varphi(u^*))-\frac{K}{\sqrt{n}} \tag{42}$$

where K depends on β,h. Now, using the same idea used in the proof of Lemma 11, we have

$$\mathbb{P}\left(\Delta\ge t\right)\le n^2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t\ge 8\beta/n$ and $\psi(L_{\max})\ge-\Delta,\psi(L_{\min})\le\Delta$ where

$$\Delta=\max_{1\le i<j\le n}\left|L_{ij}-\frac{1}{n}\sum_{k\notin\{i,j\}}\varphi(L_{ik})\varphi(L_{jk})\right|.$$

Hence there exists $\varepsilon_0>0,c>0$ such that whenever $\Delta<\varepsilon_0$ we have $L_{\min}\ge u^*-c\Delta$ and either $L_{\max}\le u^*+c\Delta$ or $|L_{\max}-v^*|\le c\sqrt{\Delta}$. Define

$$U=\{L_{\max}<(u^*+v^*)/2\}. \tag{43}$$

Then again using the idea used in Lemma 11 one can easily show that

$$\mathbb{E}(\mathbb{1}_U\cdot|L_{ij}-u^*|)\le\frac{K(\beta,h)}{n^{1/2}}\text{ for all }i<j.$$

We will show that $\mathbb{P}(U^c)\le(\log n)^2/n$ and it will imply that

$$\mathbb{E}(|L_{ij}-u^*|)\le\mathbb{E}(\mathbb{1}_U\cdot|L_{ij}-u^*|)+\mathbb{P}(U^c)\le\frac{K(\beta,h)}{n^{1/2}}\text{ for all }i<j.$$

Then the rest of the assertions follow using the steps in the proof of Theorem 23.

Hence let us concentrate on the event U^c. It is enough to restrict to the event $U^c\cap\{|L_{\max}-v^*|\le c\sqrt{\Delta}\}\cap\{L_{\min}\ge u^*-c\Delta\}$. Here the rough idea is that, a large fraction of L_{ij}'s has to be near v^* in order to make $L_{\max}\simeq v^*$. Suppose $L_{\max}=L_{i_0j_0}$. Define the set

$$A=\{k:L_{i_0k}<L_{\max}-\delta_1\}$$

where δ_1 will be chosen later such that $\delta_1+c\sqrt{\Delta}<v^*-u^*$. Note that $\varphi(u)^2\le\max\{u,u^*\}$ for all u and by assumption $|L_{\max}-v^*|\le c\sqrt{\Delta}$. Thus $\varphi(L_{ij})\le\sqrt{L_{\max}}$ for all i,j and $\varphi(L_{i_0k})\le\sqrt{L_{\max}-\delta_1}\le\sqrt{L_{\max}}(1-\delta_1/2)$ for $k\in A$. Thus we have

$$L_{\max}=L_{i_0j_0}\le\Delta+\frac{1}{n}\sum_{k\ne i_0,j_0}\varphi(L_{i_0k})\varphi(L_{j_0k})\le\Delta+L_{\max}-\frac{|A|\delta_1}{2n}$$

which clearly implies that $\frac{|A|}{n} \leq \frac{2\Delta}{\delta_1}$. Similarly define the set $A_j = \{k : L_{jk} < L_{\max} - \delta_2\}$ where δ_2 will be chosen later such that $\delta_2 + c\sqrt{\Delta} < v^* - u^*$. Using same idea as before, for $j \notin A$ we have

$$L_{\max} - \delta_1 \leq L_{i_0 j} \leq \Delta + L_{\max} - \frac{|A_j|\delta_2}{2n} \text{ or } \frac{|A_j|}{n} \leq \frac{2(\Delta+\delta_1)}{\delta_2} := M(\text{say}).$$

Choose $\delta_2 = \Delta^{1/5}, \delta_1 = \Delta^{3/5}$. Then we have

$$\begin{aligned}\sum_{i<j} |L_{ij} - L_{\max}|^2 &\leq \frac{n|A| + nM + n^2\delta_2^2}{2} \\ &\leq \frac{n^2\Delta}{\delta_1} + \frac{n^2(\Delta+\delta_1)}{\delta_2} + \frac{n^2\delta_2^2}{2} \leq 4n^2\Delta^{2/5}.\end{aligned}$$

Thus, by symmetry and Hölders' inequality, we have

$$\begin{aligned}\mathbb{E}(\mathbb{1}_{U^c} \cdot |L_{ij} - v^*|^2) \leq K\,\mathbb{E}(\mathbb{1}_{U^c} \cdot \Delta^{2/5}) &\leq K\,\mathbb{P}(U^c)^{9/10} \cdot \mathbb{E}(\Delta^4)^{1/10} \\ &\leq \frac{K(\log n)^{1/5}}{n^{1/5}}\,\mathbb{P}(U^c)^{9/10}. \qquad (44)\end{aligned}$$

for some constant K. Now using lemma 21 and equation (44) we have,

$$\begin{aligned}\mathbb{E}\left[\left|E(\mathbf{X}) - \frac{n^2\varphi(v^*)}{2}\right| \;\Big|\; U^c\right] &\leq \frac{Cn^{9/5}(\log n)^{1/5}}{\mathbb{P}(U^c)^{1/10}} \\ \text{and } \mathbb{E}\left[\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(v^*)^3}{6}\right| \;\Big|\; U^c\right] &\leq \frac{Cn^{9/5}(\log n)^{1/5}}{\mathbb{P}(U^c)^{1/10}}. \qquad (45)\end{aligned}$$

If $\mathbb{P}(U^c) > (\log n)^2/n$, from inequality (45) we have

$$\begin{aligned}\mathbb{P}\left(\left|E(\mathbf{X}) - \frac{n^2\varphi(v^*)}{2}\right| \geq Kn^{19/10} \;\Big|\; U^c\right) &\leq \frac{1}{4} \\ \text{and } \mathbb{P}\left(\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(v^*)^3}{6}\right| \geq Kn^{19/10} \;\Big|\; U^c\right) &\leq \frac{1}{4}\end{aligned}$$

for some large constant K depending on β, h. Now define the set

$$B = \left\{x \in \{0,1\}^n : \left|\frac{T(x)}{n} - \frac{n^2\varphi(v^*)^3}{6}\right| \leq Kn^{19/10}, \left|E(x) - \frac{n^2\varphi(v^*)}{2}\right| \leq Kn^{19/10}\right\}.$$

Using the same idea used in the proof of lemma 23 one can again show that

$$\left|\frac{2}{n^2}\log(Z_n\,\mathbb{P}(U^c)) - f(\varphi(v^*))\right| \leq \frac{K}{n^{1/10}}$$

for some constant K depending on β, h. The crucial fact is that $\mathbb{P}(\{L_{\max}(\mathbf{Z}) > (u^*+v^*)/2\} \cap \{\mathbf{Z} \in B\})$ is bounded away from zero when $\mathbf{Z} = ((Z_{ij}))_{i<j} \sim \mathrm{G}(n, \varphi(v^*))$. Thus we have

$$\left|\frac{2}{n^2}\log Z_n - f(\varphi(v^*))\right| \leq \frac{K}{n^{1/10}}.$$

But this leads to a contradiction, since by equation (42) we have

$$\frac{2}{n^2}\log Z_n(\beta,h) \geq f(\varphi(u^*)) - \frac{K}{\sqrt{n}}$$

and $f(\varphi(u^*)) > f(\varphi(v^*))$. Thus we have $\mathbb{P}(U^c) \leq (\log n)^2/n$ and we are done.

3.5. **Proof of Theorem 15.** The proof is almost an exact copy of the proof of Theorem 8. Recall the definition of L_{ij},

$$L_{ij} := \frac{N(\mathbf{X}^1_{(i,j)}) - N(\mathbf{X}^0_{(i,j)})}{(n-2)_{\boldsymbol{v}_F-2}} \text{ for } i<j. \tag{46}$$

In fact we can write L_{ij} explicitly as a horrible sum

$$L_{ij} = \frac{1}{\alpha_F (n-2)_{\boldsymbol{v}_F-2}} \sum_{\substack{t_1<t_2<\cdots<t_{\boldsymbol{v}_F-2}\\ t_l\in[n]\setminus\{i,j\} \text{ for all } l}} \sum_{(a,b)\in E(F)} {\sum_{\pi}}' \prod_{\substack{(k,l)\in E(F)\\ (k,l)\neq(a,b)}} X_{\pi_k\pi_l}$$

where the sum $\sum'$ is over all one-one onto map π from $V(F)=[\boldsymbol{v}_F]$ to $\{a,b,t_1,\ldots,t_{\boldsymbol{v}_F-2}\}$ where $\{\pi(a),\pi(b)\}=\{i,j\}$. Now we briefly state the main steps. First we have $\mathbb{E}(X_{ij} \mid \text{rest}) = \varphi(L_{ij})$. Moreover, using lemma 20 it is easy to see that $|\mathbb{E}(\prod_{j=1}^k X_{i_{2j-1}i_{2j}} \mid \text{rest}) - \prod_{j=1}^k \varphi(L_{i_{2j-1}i_{2j}})| \le C\beta/n$ for every distinct pairs $(i_1,i_2),\ldots,(i_{2k-1},i_{2k})$ where C is an universal constant.

Now, fix $1\le i<j\le n$. Given a configuration $\mathbf{X}$, construct another one $\mathbf{X}'$ in the following way. Choose $\boldsymbol{v}_F-2$ distinct points uniformly at random without replacement from $[n]\setminus\{i,j\}$. Replace the coordinates in $\mathbf{X}$ corresponding to the edges in the complete subgraph formed by the chosen points including i,j (except that we do not change X_{ij}) by values drawn from the conditional distribution given the rest of the edges. Call the new configuration $\mathbf{X}'$. Define the antisymmetric function $F(\mathbf{X},\mathbf{X}') := (n-2)_{\boldsymbol{v}_F-2}(L_{ij}-L'_{ij})$. and $f(\mathbf{X}) := \mathbb{E}(F(\mathbf{X},\mathbf{X}') \mid \mathbf{X})$. Using the same idea as before and Theorem 1 we have

$$\mathbb{P}\left(|L_{ij}-g_{ij}|\ge t\right) \le \exp(-cnt^2/(1+\beta)) \tag{47}$$

where c is an absolute constant and g_{ij} is obtained from L_{ij} by replacing X_{kl} by $\varphi(L_{kl})$ for all $k<l$. Note that there is a slight difference with the calculation in the triangle case, since we have to consider collections of edges where some are modified and some are not. But their contribution will be of the order of n^{-1}. Also the conditions on φ arises in the following way, if all the L_{ij}'s are constant, say equal to u, then from the "mean-field equations" for L_{ij}'s we must have

$$\begin{aligned} u &\approx \frac{1}{\alpha_F (n-2)_{\boldsymbol{v}_F-2}} \sum_{\substack{t_1<t_2<\cdots<t_{\boldsymbol{v}_F-2}\\ t_l\in[n]\setminus\{i,j\} \text{ for all } l}} \sum_{(a,b)\in E(F)} {\sum_{\pi}}' \varphi(u)^{\boldsymbol{e}_F-1} \\ &= \frac{2\boldsymbol{e}_F}{\alpha_F}\varphi(u)^{\boldsymbol{e}_F-1}. \end{aligned}$$

The next step is to show that under the conditions on φ, we have $\mathbb{E}|L_{ij}-u^*| \le Kn^{-1/2}$ for all $i<j$ where $K=K(\beta,h)$ is a constant depending only on β,h. The crucial fact is that the behavior of the function $\varphi(u)^k - au$ where $a>0$ is a positive constant and $k\ge 2$ is a fixed integer, is same as the behavior of the function $\varphi(u)^2-u$.

Now it will follow (using the same proof used for lemma 21) that

$$\begin{aligned} \mathbb{E}\left|E(\mathbf{X}) - \frac{n^2\varphi(u^*)}{2}\right| &\le Cn^{3/2} \\ \text{and } \mathbb{E}\left|N(\mathbf{X}) - \frac{(n)_{\boldsymbol{v}_F}\varphi(u^*)^{\boldsymbol{e}_F}}{\alpha_F}\right| &\le Cn^{\boldsymbol{v}_F-1/2} \end{aligned}$$

where C is a constant depending only on β, h. The rest of the proof follows using the arguments used in the proof of Theorem 8.

Proof of Theorem 14. Using the method of proof for the triangle case and the result from Theorem 15 the proof follows easily. $\square$

Proof of Lemma 16. The proof is same as the proof of lemma 12 except for the constants. $\square$

3.6. **Proof of Theorem 17.** Suppose $\boldsymbol{\sigma}$ is drawn from the Gibbs distribution $\mu_{\beta,h}$. We construct $\boldsymbol{\sigma'}$ by taking one step in the heat-bath Glauber dynamics as follows: Choose a position I uniformly at random from Ω, and replace the I-th coordinate of $\boldsymbol{\sigma}$ by an element drawn from the conditional distribution of the σ_I given the rest. It is easy to see that $(\boldsymbol{\sigma}, \boldsymbol{\sigma'})$ is an exchangeable pair. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma'}) := |\Omega|(m(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma'})) = \sigma_I - \sigma'_I$$

be an antisymmetric function in $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$. Since the Hamiltonian is a simple explicit function, one can easily calculate the conditional distribution of the spin of the particle at position x given the spins of the rest. In fact we have $\mathbb{E}(\sigma_x|\{\sigma_y, y \neq x\}]) = \tanh(2\beta d m_x(\boldsymbol{\sigma}))$ where $m_x(\boldsymbol{\sigma}) := \frac{1}{2d}\sum_{y\in N_x} \sigma_y$ is the average spin of the neighbors of x for $x \in \Omega$. Now using Fourier-Walsh expansion we can write the function $\tanh(2\beta d m_x(\boldsymbol{\sigma}))$ as sums of products of spins in the following way. We have

$$\tanh(2d\beta m_x(\boldsymbol{\sigma})) = \sum_{k=0}^{2d} a_k(\beta) \sum_{|S|=k, S\subseteq N_x} \sigma_S \tag{48}$$

where

$$a_k(\beta) := \frac{1}{2^{2d}} \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}} \tanh\left(\beta\sum_{i=1}^{2d}\sigma_i\right)\prod_{j=1}^{k}\sigma_j \tag{49}$$

for $k = 0, 1, \ldots, 2d$. It is easy to see that $a_k(\beta) = 0$ if k is even and $a_k(\beta)$ is a rational function of $\tanh(2\beta)$ if k is odd. Note that the dependence of a_k on d is not stated explicitly. Thus using equation (48) and the definitions in (20) we have

$$\begin{aligned}
f(\boldsymbol{\sigma}) = \mathbb{E}[F(\boldsymbol{\sigma}, \boldsymbol{\sigma'})|\boldsymbol{\sigma}] &= \frac{1}{|\Omega|}\sum_{x\in\Omega} E[\sigma_x - \sigma'_x|\boldsymbol{\sigma}] \\
&= m(\boldsymbol{\sigma}) - \frac{1}{|\Omega|}\sum_{x\in\Omega}\tanh(2\beta d m_x(\boldsymbol{\sigma})) \\
&= (1 - 2da_1(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\binom{2d}{2k+1} a_{2k+1}(\beta) r_{2k+1}(\boldsymbol{\sigma}).
\end{aligned}$$

Define $\theta_k(\beta) := \binom{2d}{2k+1}a_{2k+1}(\beta)$ for $k = 0, 1, \ldots, d-1$. Note that we can explicitly calculate the value of $\theta_0(\beta)$ as follows,

$$\theta_0(\beta) = \frac{1}{4^d}\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}} \tanh\left(\beta\sum_{i=1}^{2d}\sigma_i\right)\sum_{i=1}^{2d}\sigma_i = \frac{2}{4^d}\sum_{k=1}^{d} 2k\binom{2d}{d+k}\tanh(2k\beta).$$

Now we have $|F(\boldsymbol{\sigma},\boldsymbol{\sigma}')| \le 2$ and

$$|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \le \frac{2}{|\Omega|}\left(|1-\theta_0(\beta)| + \sum_{k=1}^{d-1}(2k+1)\theta_k(\beta)\right) = \frac{2}{|\Omega|}b(\beta)$$

for all values of $\boldsymbol{\sigma},\boldsymbol{\sigma}'$. Hence the condition of Theorem 1 is satisfied with $B = 0, C = 2|\Omega|^{-1}b(\beta)$. So by part (ii) of Theorem 1 we have

$$\mathbb{P}\left(\sqrt{|\Omega|}\left|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma})\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{4b(\beta)}\right)$$

for all $t > 0$. Obviously $\theta_0(\cdot)$ is a strictly increasing function of β. Also we have $\theta_0(0) = 0$ and

$$\theta_0(\infty) := \lim_{\beta\to\infty}\theta_0(\beta) = \frac{1}{4^{d-1}}\sum_{k=1}^{d} k\binom{2d}{d+k}.$$

For $d = 1$ we have $\theta_0(\infty) = 1$ and for $d \ge 2$ we have

$$\begin{aligned}\theta_0(\infty) &\ge \frac{1}{4^{d-1}}\left[2\sum_{k=1}^{d}\binom{2d}{d+k} - \binom{2d}{d+1}\right]\\ &= \frac{1}{4^{d-1}}\left[2^{2d} - \binom{2d}{d} - \binom{2d}{d+1}\right] = 4 - \frac{8}{2^{2d+1}}\binom{2d+1}{d+1}\end{aligned}$$

and from the fact that $\sum_{k=d-1}^{d+2}\binom{2d+1}{k} \le 2^{2d+1}$ we have

$$\frac{1}{2^{2d+1}}\binom{2d+1}{d+1} \le \frac{d+2}{4(d+1)} \le \frac{1}{3} \text{ for } d \ge 2.$$

Hence for $d \ge 2$ we have $\theta_0(\infty) > 1$ and there exists $\beta_1 \in (0,\infty)$, depending on d, such that $1 - \theta_0(\beta) > 0$ for $\beta < \beta_1$ and $1 - \theta_0(\beta) < 0$ for $\beta > \beta_1$. This completes the proof.

3.7. **Proof of Proposition 19.** The proof is almost same as the proof of proposition 17. Define $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ as before. Define the antisymmetric function $F(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ as follows

$$\begin{aligned}F(\boldsymbol{\sigma},\boldsymbol{\sigma}') &:= |\Omega|(1+\tanh(h)\tanh(2\beta d m_I(\boldsymbol{\sigma})))(m(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma}'))\\ &= (1+\tanh(h)\tanh(2\beta d m_I(\boldsymbol{\sigma})))(\sigma_I - \sigma_I').\end{aligned}$$

Recall that $m_x(\boldsymbol{\sigma}) := \frac{1}{2d}\sum_{y\in N_x}\sigma_y$ is the average spin of the neighbors of x for $x \in \Omega$. Now under $\mu_{\beta,h}$ we have

$$\begin{aligned}\mathbb{E}(\sigma_x|\{\sigma_y, y \ne x\}) &= \tanh(2\beta d m_x(\boldsymbol{\sigma}) + h)\\ &= \frac{\tanh(h) + \tanh(2\beta d m_x(\boldsymbol{\sigma}))}{1+\tanh(h)\tanh(2\beta d m_x(\boldsymbol{\sigma}))}.\end{aligned}$$

Thus we have

$$\begin{aligned}f(\boldsymbol{\sigma}) &= \mathbb{E}(F(\boldsymbol{\sigma},\boldsymbol{\sigma}')|\boldsymbol{\sigma})\\ &= \frac{1}{|\Omega|}\sum_{x\in\Omega}(1+\tanh(h)\tanh(2\beta d m_x(\boldsymbol{\sigma})))\,\mathbb{E}(\sigma_x - \sigma_x'|\boldsymbol{\sigma})\\ &= m(\boldsymbol{\sigma}) - \tanh(h) + \frac{1}{|\Omega|}\sum_{x\in\Omega}(\tanh(h)\sigma_x - 1)\tanh(2\beta d m_x(\boldsymbol{\sigma})).\end{aligned}$$

After some simplifications and using the definitions of the functions r, s we have

$$\begin{aligned} f(\boldsymbol{\sigma}) = (1-\theta_0(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma}) \\ - \tanh(h)\left(1-\sum_{k=0}^{d-1}\theta_k(\beta)s_{2k+1}(\boldsymbol{\sigma})\right). \end{aligned}$$

Now for all values of $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ we have

$$|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \le \frac{2}{|\Omega|}b(\beta)(1+\tanh|h|)$$

and the proof onwards is exactly as in the proof of proposition 17.

3.8. **Proof of Theorem 2.** Assume that $\psi(0) > 0$. We will handle the case $\psi(0) = 0$ later. Note that condition (1) implies that $x^\alpha/\psi(x)$ is a nondecreasing function for $x > 0$. Define the function

$$\varphi(x) := \frac{x^2}{\psi(x)} \text{ and } \gamma(x) := 2 - \frac{x\psi'(x)}{\psi(x)} \text{ for } x \neq 0$$

and $\varphi(0) = 0, \gamma(0) = 2$. Clearly we have $2-\alpha \le \gamma(x) \le 2$ for all $x \in \mathbb{R}$. Now, $\limsup_{x\to 0}\varphi(x) \le \lim_{x\to 0+} x^{2-\alpha}/\psi(1) = 0 = \varphi(0)$ as $\alpha < 2$. Also $\varphi(x)$ is differentiable in $\mathbb{R}\setminus\{0\}$ with

$$\varphi'(x) = \frac{x\gamma(x)}{\psi(x)} > 0 \text{ for } x \neq 0. \tag{50}$$

Hence φ is absolutely continuous in $\mathbb{R}$ and is increasing for $x \ge 0$.

Define $Y = f(X)$. First we will prove that all moments of $\varphi(Y)$ are finite. Next we will estimate the moments which will in turn show that $\varphi(Y)^{1/2}$ has finite exponential moment in $\mathbb{R}$. Finally using Chebyshev's inequality we will prove the tail probability.

By monotonicity of ψ in $[0,\infty)$ and definition of α, we have

$$0 \le \frac{x\psi'(x)}{\psi(x)} \le \alpha \text{ for all } x \ge 0. \tag{51}$$

It also follows from (50) that $0 \le (\log\varphi(x))' \le 2/x$ for $x > 0$ and integrating we have $\varphi(x) \le \varphi(1)x^2$ for all $x \ge 1$. Hence $\varphi(x) = \varphi(|x|) \le \varphi(1)(1+x^2)$ for all $x \in \mathbb{R}$ and this, combined with our assumption that $\mathbb{E}(|f(X)|^k) < \infty$ for all $k \ge 1$, implies that $\mathbb{E}(\varphi(Y)^k) < \infty$ for all $k \ge 1$.

Define

$$\beta := \left\lceil \frac{5(2-\alpha)+\delta+1/4}{(2-\alpha)^2} \right\rceil \ge 3.$$

Fix an integer $k \ge \beta$ and define

$$g(x) = \frac{x^{2k-1}}{\psi^k(x)} \text{ and } h(x) = \frac{x^{2k-2}}{\psi^k(x)} \text{ for } x \in \mathbb{R}.$$

Clearly $\mathbb{E}(|Yg(Y)|) < \infty$. Note that g, h are continuously differentiable in $\mathbb{R}$ as $k \ge 3$. Moreover, for $x \in \mathbb{R}$ we have, $|g'(x)| = h(x)\,|k\gamma(x)-1| \le (2k-1)h(x)$, $h'(x) = (k\gamma(x)-2)\,x^{2k-3}/\psi^k(x)$ and

$$h''(x) = \left[(k\gamma(x)-2)\,(k\gamma(x)-3) + kx\gamma'(x)\right]\frac{x^{2k-4}}{\psi^k(x)}.$$

We also have

$$x\gamma'(x) = -\frac{x\psi'(x)}{\psi(x)}\left(1 - \frac{x\psi'(x)}{\psi(x)}\right) - \frac{x\psi''(x)}{\psi(x)} \geq -1/4 - \delta$$

for $x \in \mathbb{R}$. Now $k \geq \beta$ implies that

$$(k\gamma(x) - 2)\,(k\gamma(x) - 3) + kx\gamma'(x) \geq (k(2-\alpha) - 2)\,(k(2-\alpha) - 3) - k(\delta + 1/4) \geq 0$$

for all x. Thus $h''(x) \geq 0$ for all x and h is convex in $\mathbb{R}$.

Let $X', F(X, X')$ be as given in the hypothesis. Define $Y' = f(X')$. Recall that (X, X') is an exchangeable pair and so is (Y, Y'). Using the fact that $f(X) = \mathbb{E}(F(X, X')|X)$ almost surely, exchangeability of (X, X') and antisymmetry of F, we have

$$\begin{aligned}\mathbb{E}(Yg(Y)) = \mathbb{E}(f(X)g(Y)) &= \mathbb{E}(F(X, X')g(Y)) \\ &= \frac{1}{2}\,\mathbb{E}(F(X, X')(g(Y) - g(Y'))).\end{aligned} \tag{52}$$

Now, for any $x < y$ we have

$$\left|\frac{g(x) - g(y)}{x - y}\right| = \left|\int_0^1 g'(tx + (1-t)y)\,dt\right| \leq (2k-1)\int_0^1 h(tx + (1-t)y)dt$$

and convexity of h implies that

$$\int_0^1 h(tx + (1-t)y)dt \leq \int_0^1 (th(x) + (1-t)h(y))dt = (h(x) + h(y))/2.$$

Hence, from equation (52) we have

$$\begin{aligned}\mathbb{E}(Yg(Y)) &\leq \frac{2k-1}{4}\,\mathbb{E}(|(Y - Y')F(X, X')|(h(Y) + h(Y'))) \\ &= (2k-1)\,\mathbb{E}(\Delta(X)h(Y)) \leq (2k-1)\,\mathbb{E}(\psi(Y)h(Y))\end{aligned} \tag{53}$$

where the equality follows by definition of $\Delta(X)$ and exchangeability of (Y, Y'). Thus for any $k \geq \beta$ we have, from (53),

$$\mathbb{E}(\varphi(Y)^k) \leq (2k-1)\,\mathbb{E}(\varphi(Y)^{k-1}). \tag{54}$$

Using induction for $k \geq \beta$ we have

$$\mathbb{E}(\varphi(Y)^k) \leq \frac{(2k)!2^\beta\beta!}{2^k k!(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta) \text{ for } k \geq \beta.$$

Also Hölder's inequality applied to (54) for $k = \beta$ implies that $\mathbb{E}(\varphi(Y)^\beta) \leq (2\beta - 1)^\beta$. Thus we have

$$\mathbb{E}(\varphi(Y)^k) \leq \begin{cases} \frac{(2k)!2^\beta\beta!}{k!2^k(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta) & \text{if } k > \beta \\ (2\beta-1)^k & \text{if } 0 \leq k \leq \beta. \end{cases} \tag{55}$$

Note that we have $e^x \leq e^x + e^{-x} = 2\sum_{k\geq 0} x^{2k}/(2k)!$ for all $x \in \mathbb{R}$. Combining everything we finally have

$$\begin{aligned}\mathbb{E}(\exp(\theta\varphi(Y)^{1/2})) &\leq 2\sum_{k=0}^{\infty}\frac{\theta^{2k}}{(2k)!}\,\mathbb{E}(\varphi(Y)^k) \\ &\leq \frac{2^{\beta+1}\beta!}{(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta)\sum_{k=\beta}^{\infty}\frac{\theta^{2k}}{2^k k!} + \sum_{k=0}^{\beta-1}\frac{2(2\beta-1)^k\theta^{2k}}{(2k)!} \leq C_\beta\exp(\theta^2/2)\end{aligned}$$

for all $\theta \geq 0$ where the constant C_β is given by

$$C_\beta := \max\left\{ \frac{2(2\beta-1)^k 2^k k!}{(2k)!} \middle| 0 \leq k \leq \beta \right\}.$$

Here we used the fact that $(2k)! \geq 2^{2k-1}k!^2/k$. Now recall that φ is an increasing function in $[0,\infty)$. Thus using Chebyshev's inequality for $\exp(\theta\varphi(x)^{1/2})$ with $\theta = \varphi(t)^{1/2}$ we have

$$\mathbb{P}(|f(X)| \geq t) \leq C_\beta e^{-\theta\varphi(t)^{1/2}+\theta^2/2} = C_\beta e^{-\varphi(t)/2}.$$

Now suppose that $\psi(0) = 0$. For $\varepsilon > 0$ fixed, define $\psi_\varepsilon(x) = \psi(x) + \varepsilon$. Clearly we have $\Delta(X) \leq \psi_\varepsilon(f(X))$ a.s. and ψ_ε satisfies all the other properties of ψ including

$$x\psi_\varepsilon'(x)/\psi_\varepsilon(x) = x\psi_\varepsilon'(x)/\psi(x) \cdot \psi(x)/(\psi(x)+\varepsilon) \leq \alpha$$
$$\text{and } x\psi_\varepsilon''(x)/\psi_\varepsilon(x) = x\psi_\varepsilon''(x)/\psi(x) \cdot \psi(x)/(\psi(x)+\varepsilon) \leq \delta$$

for all $x > 0$. Hence all the above results hold for ψ_ε and $\varphi_\varepsilon(x) = x^2/\psi_\varepsilon(x)$. Now $\varphi_\varepsilon \uparrow \varphi$ as $\varepsilon \downarrow 0$. Letting $\varepsilon \downarrow 0$ we have the result.

When ψ is once differentiable with $\alpha < 2$, it is easy to see that the function h is nondecreasing (need not be convex) in $[0,\infty)$ for $k \geq \beta := \lceil 2/(2-\alpha) \rceil$. In that case we have

$$\int_0^1 h(tx + (1-t)y)dy \leq \max_{z\in[x,y]} h(z) \leq h(x) + h(y)$$

for $x \leq y$. Hence we have the recursion

$$\mathbb{E}(\varphi(Y)^k) \leq 2(2k-1)\,\mathbb{E}(\varphi(Y)^{k-1}) \tag{56}$$

for $k \geq \beta$. Using the same proof as before it then follows that

$$\mathbb{P}(|f(X)| \geq t) \leq Ce^{-\varphi(t)/4}$$

where C depends only on α. $\square$

Acknowledgments. The authors thank Amir Dembo, Erwin Bolthausen, Ofer Zeitouni and Persi Diaconis for various helpful discussions and comments. They would also like to thank an anonymous referee for helpful comments, pointing out several typos and encouraging them to write down Theorem 2 in the most general form.

References

[1] Franck Barthe, Patrick Cattiaux, and Cyril Roberto, *Concentration for independent random variables with heavy tails*, AMRX Appl. Math. Res. Express **2** (2005), 39–60.

[2] Franck Barthe, Patrick Cattiaux, and Cyril Roberto, *Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry*, Rev. Mat. Iberoam. **22** (2006), no. 3, 993–1067.

[3] S. Bhamidi, G. Bresler, and A. Sly, *Mixing time of exponential random graphs*, Proc. of the 49th Annual IEEE Symp. on FOCS (IEEE Computer Society Washington, DC, USA, 2008), pp. 803–812.

[4] S. G. Bobkov, *Large deviations and isoperimetry over convex probability measures with heavy tails*, Electron. J. Probab. **12** (2007), 1072–1100 (electronic).

[5] S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal. **10** (2000), no. 5, 1028–1052.

[6] Béla Bollobás, *Random graphs*, Second, Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[7] E. Bolthausen, *Laplace approximations for sums of independent random vectors. II. Degenerate maxima and manifolds of maxima*, Probab. Theory Related Fields **76** (1987), no. 2, 167–206.

[8] E. Bolthausen, F. Comets, and A. Dembo, *Large deviations for random matrices and random graphs.* (2009), In preparation.

[9] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities using the entropy method*, Ann. Probab. **31** (2003), no. 3, 1583–1614.

[10] Sourav Chatterjee, *Concentration inequalities with exchangeable pairs.*, Ph.D. Theis, Department of Statistics, Stanford University, 2005, Available at `http://arxiv.org/abs/math.PR/0507526`.

[11] Sourav Chatterjee, *Stein's method for concentration inequalities*, Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321.

[12] Sourav Chatterjee, *Concentration of Haar measures, with an application to random matrices*, J. Funct. Anal. **245** (2007), no. 2, 379–389.

[13] Sourav Chatterjee and Q.-M. Shao, *Stein's Method of Exchangeable Pairs with Application to the Curie-Weiss Model* (2009), Available at `http://arxiv.org/abs/0907.4450`.

[14] J.-R. Chazottes, P. Collet, C. Külske, and F. Redig, *Concentration inequalities for random fields via coupling*, Probab. Theory Related Fields **137** (2007), no. 1-2, 201–225.

[15] H. Doring and P. Eichelsbacher, *Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices* (2009), Preprint, Available at `http://arxiv.org/abs/0901.3246`.

[16] P. Eichelsbacher and M. Lowe, *Stein's method for dependent random variables occurring in Statistical Mechanics* (2009), Preprint, Available at `http://arxiv.org/abs/0908.1909`.

[17] Richard S. Ellis and Charles M. Newman, *The statistics of Curie-Weiss models*, J. Statist. Phys. **19** (1978), no. 2, 149–161.

[18] Richard S. Ellis and Charles M. Newman, *Limit theorems for sums of dependent random variables occurring in statistical mechanics*, Z. Wahrsch. Verw. Gebiete **44** (1978), no. 2, 117–139.

[19] Richard S. Ellis, *Entropy, large deviations, and statistical mechanics*, Grundlehren der Mathematischen Wissenschaften, vol. 271, Springer-Verlag, New York, 1985.

[20] Ivan Gentil, Arnaud Guillin, and Laurent Miclo, *Modified logarithmic Sobolev inequalities and transportation inequalities*, Probab. Theory Related Fields **133** (2005), no. 3, 409–436.

[21] Nathael Gozlan, *Characterization of Talagrand's like transportation-cost inequalities on the real line*, J. Funct. Anal. **250** (2007), no. 2, 400–425.

[22] Nathael Gozlan, *Poincare inequalities and dimension free concentration of measure*, To appear in Ann. Inst. Henri Poincaré Probab. Stat. (2009).

[23] Barry Simon and Robert B. Griffiths, *The* $(\phi^4)_2$ *field theory as a classical Ising model*, Comm. Math. Phys. **33** (1973), 145–164.

[24] E. Ising, *Beitrag zur theorie des ferromagnetismus*, Zeitschrift für Physik A Hadrons and Nuclei **31** (1925), no. 1, 253–258.

[25] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[26] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142** (2004), 61–92.

[27] Svante Janson and Andrzej Ruciński, *The infamous upper tail*, Random Structures Algorithms **20** (2002), no. 3, 317–342. Probabilistic methods in combinatorial optimization.

[28] J. H. Kim and V. H. Vu, *Divide and conquer martingales and the number of triangles in a random graph*, Random Structures Algorithms **24** (2004), no. 2, 166–174.

[29] R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric aspects of functional analysis, 2000, pp. 147–168.

[30] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.

[31] Anders Martin-Löf, *A Laplace approximation for sums of independent random variables*, Z. Wahrsch. Verw. Gebiete **59** (1982), no. 1, 101–115.

[32] Lars Onsager, *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*, Phys. Rev. **65** (1944), no. 3-4, 117–149.

[33] Juyong Park and M. E. J. Newman, *Statistical mechanics of networks*, Phys. Rev. E **70** (2004), no. 6, 066117.

[34] Juyong Park and M. E. J. Newman, *Solution for the properties of a clustered network*, Phys. Rev. E **72** (2005), no. 2, 026136.

[35] Martin Raič, *CLT-related large deviation bounds based on Stein's method*, Adv. in Appl. Probab. **39** (2007), no. 3, 731–752.

[36] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, 1972, pp. 583–602.

[37] Charles Stein, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986.

[38] Michel Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. **81** (1995), 73–205.

[39] Van H. Vu, *A large deviation result on the number of small subgraphs of a random graph*, Combin. Probab. Comput. **10** (2001), no. 1, 79–94.

Sourav Chatterjee
Department of Statistics, University of California Berkeley
367 Evans Hall, Berkeley, CA 94720-3860, USA.
Email: `sourav@stat.Berkeley.EDU`
URL: `http://www.stat.berkeley.edu/~sourav`

Partha Sarathi Dey
Department of Statistics, University of California Berkeley
367 Evans Hall, Berkeley, CA 94720-3860, USA.
Email: `partha@stat.Berkeley.EDU`
URL: `http://www.stat.berkeley.edu/~partha`